Three Essays on the Economic Theory of Self-Control

For the Degree of PhD by Research

Alla Doubrovina

The University of Edinburgh

22 September 2009

Contents

Declaration of Own Work

I hereby declare that this thesis is the result of my own work and does not contain any work done in collaboration with others, except as stated in the acknowledgements. References to the work of other people are explicitly stated in the text. I further declare that no part of this thesis has been submitted for any other degree or qualification at any other university.

Alla Doubrovina

Edinburgh University

September 2009

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Ed Hopkins, for his advice, guidance and encouragement over my four years at Edinburgh. Ed has always willingly made time to listen to my thoughts and logic, offer constructive comments and read sometimes last minute drafts of this thesis. I would also like to thank Ed for the trust he put in me at the very beginning of my graduate studies and for the friendship that I hope will last beyond this work.

I would also like to thank Professor József Sákovics for kindly offering his help whilst Ed was on sabbatical and for his meticulous advice on the many versions of my chapters ever since.

Summary

The prevalence of obesity in the developed world has more than doubled over the last 30 years with disproportionate burden of weight falling on the lower income groups. In our first chapter, we argue that advances in food technology, which have made food abundant, available and cheap, have also made obesity a problem of self-control. We then develop a model of self-control which can explain the inverse socioeconomic gradient of obesity. We argue that whereas the payoffs to simple temptations of palatable foods are independent of income, people's underlying wellbeing, such as the utility of their everyday 'sensible' life, is increasing in income. This implies that the relative payoffs to the temptations of energy-rich, palatable food, compared to the underlying wellbeing, may be higher for lower income groups, which reduces their incentive to exercise self-control, and leads to higher rates of obesity. Our model is based on the premise that people can make systematic mistakes, which we represent through introduction of a second myopic self. Further, the propensity for systematic mistakes is endogenously determined. Based on evidence from psychology, we argue that exercise of self-control is costly in the short run, but increases the stock of willpower and so reduces the propensity for systematic mistakes in the long run. Incorporating the short-term vs. long term dynamics of willpower and allowing individuals to endogenously affect their self-control ability is the secondary objective of the chapter.

In the second chapter, we develop a model of job autonomy, human capital and self-control which aims to explain the effect of different types of occupations on self-control outcomes, which is distinct from the pure income effect of wages. Jobs differ in the degree of autonomy placed on the worker. We argue that successful performance in autonomous jobs requires the kind of human capital, acquiring which demands exercise of self-control in the first place. Accumulating such human capital then has spillover effects on individual's level of willpower in other areas of his life. We show that an increase in the degree of job autonomy in fact increases the steady state levels of willpower, self-control and human capital. Increasing the return to human capital has a similar effect. We also find an upper bound for marginal cost of self-control for which a small increase in autonomy increases agents' experienced welfare in steady state.

In the third chapter, we re-visit the hyperbolic discounting view of self-control by extending the Benabou and Tirole's "Willpower and Personal Rules", (2004), model to explore intergenerational links in self-control outcomes. Benabou and Tirole build a self-signaling model of personal rules based on self-reputation, in which people are uncertain about their underlying willpower type but can infer it from their own past actions. However, in equilibrium of their model, full spectrum of self-control outcomes can be achieved depending on agents' initial beliefs, which remain exogenous. In this chapter, we put their self-signaling model in the dynamic overlapping generations context, which

provides a mechanism for the formation of initial beliefs and generates heterogeneous behaviour among agents of the same type driven by different parental choices. We show that, conditional on type, children of parents who exercised more self-control during their lifetime, have higher self-confidence, exercise more self-control themselves and are at least ex ante better off. We find that this heterogeneity persists from two to infinite generations set-up with the long run fraction of population exercising self-control being lower with the influence of parental behaviour than without. Introduction of parental altruism retains the heterogeneity of children's behaviour but also induces parents to exercise more self-control, especially when observed by children in later stages of their life.

Part 1

Why Obesity Weighs on the Poor: A Model of Self-Control

Abstract

The prevalence of obesity in the developed world has more than doubled over the last 30 years with disproportionate burden of weight falling on the lower income groups. In our first chapter, we argue that advances in food technology, which have made food abundant, available and cheap, have also made obesity a problem of self-control. We then develop a model of self-control which can explain the inverse socioeconomic gradient of obesity. We argue that whereas the payoffs to simple temptations of palatable foods are independent of income, people's underlying wellbeing, such as the utility of their everyday 'sensible' life, is increasing in income. This implies that the relative payoffs to the temptations of energy-rich, palatable food, compared to the underlying wellbeing, may be higher for lower income groups, which reduces their incentive to exercise self-control, and leads to higher rates of obesity. Our model is based on the premise that people can make systematic mistakes, which we represent through introduction of a second myopic self. Further, the propensity for systematic mistakes is endogenously determined. Based on evidence from psychology, we argue that exercise of self-control is costly in the short run, but increases the stock of willpower and so reduces the propensity for systematic mistakes in the long run. Incorporating the short-term vs. long term dynamics of willpower and allowing individuals to endogenously affect their self-control ability is the secondary objective of the chapter.

1 Introduction

One billion people in the world today are estimated to be overweight, with at least 300 million classified as clinically obese (WHO, 2004). Thirty years ago, these numbers were almost three times lower. The dramatic rise in body weights since the 1980s, particularly in the developed world, is in sharp contrast to the gentle increase in average weights which occurred throughout the rest of the 20th century. A hundred years ago, weights were still largely below the longevity optimum, as the developed countries had only recently shaken off subsistence poverty, malnutrition and communicable diseases. The rising body weights at the time were a major source of better health (Fogel, 1994); in fact, improving the nutrition of the working class was even seen as an important part of the growth strategy[1]. However, since the 1960s in the USA and for the last three decades in Britain and many other European countries, weights have been rising above the optimal longevity norm, with the effect of rising weight becoming detrimental rather than beneficial to health. From the late 1970s to the early 2000s, the prevalence of obesity in the USA doubled to a level of 30% (Flegal et al, 2002). In Britain, obesity increased nearly three-fold over the same period to a level of 22.7% (HSE, 2006). This rise in obesity has been accompanied by a significant rise in obesity-related diseases, such as type-2 diabetes, cardiovascular disease, several types of cancer, musculoskeletal disorders and gallbladder disease. Large longitudinal studies in the US have shown that obesity doubles mortality risk (Bray et al, 2008). Yet the distribution of obesity and its associated health and medical costs is not uniform across the society: obesity follows an inverse socioeconomic gradient with the heaviest burden falling onto the least privileged groups.

In this paper, we develop a model of self-control which can explain the socioeconomic gradient of obesity. We argue that advances in food technology, which have made food available, easily-accessible, ready-to-eat and cheap, have also made obesity a problem of self-control. Developed in the times of food insecurity, the human genotype is maladapted to the conditions of permanent food abundance[2]; in the current 'obesogenic environment'[3], maintaining a desirable weight requires a conscious and costly effort. We call this effort self-control. We argue further that whereas the payoffs to simple temptations, such as indulgence in palatable foods, are independent of income, people's underlying well-being, such as their enjoyment of a 'sensible' way of life, is increasing with income[4]. This makes the payoffs to the temptations of energy-rich palatable foods *relative* to the utility of the rest of their everyday 'sensible' lives higher among the poor, reducing their incentive to put in the self-control effort required to maintain a slender waistline, and increasing the incentives for weight gain. Thus, our model predicts that lower income can lead to lower self-control, based on the simple assumption that utility is increasing in income. This effect is amplified if higher income raises the marginal return to willpower, but this extra condition is not required to drive our basic result. We also show how an increase in food abundance can lead to a reduction in self-control across income groups.

Our model is based on the premise that people can make systematic mistakes; reducing one's propensity for such mistakes is at the heart of self-control. To represent

[1] The Food and Agriculture Organisation in the US was created to increase availability of low cost calorie sources to the working class in the first half of the 20th century.

[2] Ulijaszek, 2007; Neel at al, 1998

[3] A term coined by Swinburn et al. (1999) who argue that the physical, economic, social and cultural environments of the majority of the developed countries encourage a positive energy balance, whereby the energy value of food consumed exceeds the energy expenditure levels.

[4] Higher incomes can give people the opportunity to derive utility from 'sensible' goods and activities which may not be accessible to the lower income groups.

the propensity for systematic mistakes, we undertake a departure from standard intertemporal choice models by introducing a myopic second self as a counterpart to the otherwise rational agent. The rational self is sophisticated in the sense that he understands the process through which the myopic self might come to make decisions, and acts to maximise lifetime utility, taking the myopic self into account. Our 'systematic mistake' interpretation of the second self is in direct parallel to Bernheim and Rangel (2004). In terms of mathematics, decisions of the rational self correspond to a solution of an infinite continuous time dynamic programming problem with no uncertainty.

The propensity for systematic mistakes is endogenously determined in our model. Based on evidence from experimental psychology, we believe that, in addition to being depleted in the short run, willpower is strengthened in the long run through repeated use. In terms of short-term effects, experiments by Muraven (1999), Baumeister et al (1998, 2003, 2005), Vohs and Faber (2007), Vohs and Heatherton (2004) have extensively documented a pattern of 'ego depletion', where exercise of self-control on an initial task impairs the subjects' ability for self-regulation on a subsequent and unrelated task. For the long run effects, the most conclusive evidence that repeated exercise of self control leads to an improvement in the self-control ability in the long run comes from a series of experiments by Oaten and Cheng (2006, 2007), in which they show that a two or four months adherence to an exercise regime or financial monitoring scheme can improve subjects' performance not only in the experimental self-control tasks, but also in terms of healthier eating and better studying habits. Thus, we assign a convex cost to self-control in the short run, but also allow the stock of willpower to accumulate with exercise of self-control over time. The probability of the myopic self making a decision, i.e. the probability of a mistake, then declines with willpower. Incorporating the short-run vs. long-run dynamic properties of willpower, and thereby endowing the individual with some ability to control his alter ego, is the secondary objective of this paper.

Although we rely on the dual self terminology, our approach to self-control is quite different from most of the dual self literature. The focus of this literature is on the interaction of two distinct systems or selves with distinct preferences or objectives, where one system can implicitly control the other. In Fudenberg and Levine (2006), it is the long run self who, at some cost to utility, has the capacity to alter the preferences and hence behaviour of the successive short run selves. In Benhabib and Bisin (2005), it is the supervisory function which can use the goal-based controlled processes to override the initial (emotion-based) response of automatic processes, and is activated if the consequences of the automatic decision are too costly. In Loewenstein and O'Donoghue (2007), behaviour is the result of the interaction of deliberative and affective processes, where both processes can actually influence each other. The fact that one self or system can alter or 'override' behaviour of the other means that self-control is always possible in these models, albeit at some cost.

We take an alternative approach: self-control is not always possible in our model, as the individual will continue to make some mistakes; nor can the rational self alter the behaviour of the myopic self. What the rational self can do is control the frequency with which myopic decisions are made. This amounts to self-control in our interpretation. Whether it is the frequency or the nature of the second self/system decisions that the rational self/system should be able to influence can be subject to debate. Perhaps it should be both. However, what our approach allows us to do is to incorporate the intuitive and psychologically supported fact that willpower, although costly in the short-run, can grow over time. Thus, we allow the individual to get better at self-control.

We are aware of one other paper to date that explicitly models willpower as a limited resource and gives some consideration to the short-run vs. long-run trade-offs. Ozdenoren, Salant and Silverman (2006) consider a canonical single-agent finite cake-eating problem with the addition of willpower constraint. In their model, moderating consumption requires willpower but willpower also has an alternative use. This alternative use drives the main the result of model, which lies in generating upward-sloping consumption paths. Whereas this result is both novel and interesting, the brief reference the authors make to the link between wealth and willpower is less convincing. In particular, they find that two agents who differ only in the size of the initial cake will have the same absolute level of consumption, but the poorer agent runs out of the cake sooner and so appears less disciplined. If the initial endowment of the richer agent is sufficiently large, he would then also have more willpower left for the alternative activity. We think our intuition, which relies on the difference in relative payoffs of temptation vs 'sensible' life and does not produce the same consumption between rich and poor, is more plausible.

2 Obesity: A Rising Problem of Self-Control

> "Obesity shows how abundance, through cheapness, variety, novelty, and choice, could make a mockery of the rational consumer, how it enticed only in order to humiliate". Avner Offer

The first notable fact behind the obesity problem of the developed world today is its widespread prevalence: 47.5% of people in Europe and 64% in the USA are now classified as overweight or obese (HSE, 2006 and Flegal et al., 2002). Globally, about 300mln people are estimated to be obese (International Obesity Taskforce, 2004). The second fact refers to the rate at which obesity has been spreading. In the 2006 Health Survey for England, 24% of adults were classified as obese. Ten years earlier, the figure was only 17.5% (Fig.1a). The third fact is that obesity follows an inverse socioeconomic gradient, with the worst burden falling on the least privileged social groups and ethnic minorities. Already in 1989, Sobal and Stunkard reviewed 144 studies on the prevalence of obesity and found overwhelming evidence for the inverse association of socioeconomic status and obesity in the developed countries, particularly among women. Their review has been updated since by Ball and Crawford (2005) and McLaren (2007) with largely the same findings. Specifically, in their Healthy People 2010 Report, the US Department of Health and Human Services found that in 2002, higher income level households registered 29% of obesity whereas the low income level households registered 36%[5]. The 2006 Health Survey for England showed similar trends (Fig.1b); women in particular registered a 13% difference in obesity rates between the highest and the lowest income quintiles[6].

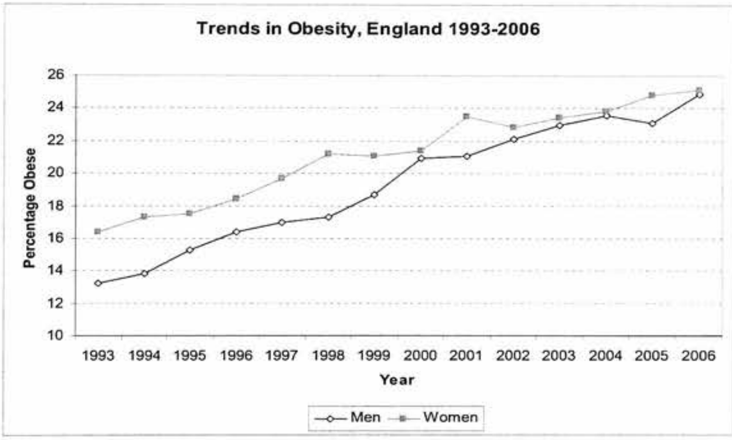


Figure 1a. Rising rates of obesity in England, 1993-2006

[5]In the US DHSS Healthy People 2010 report, households are classified as 'higher income level' if their income is greater than 130% of poverty threshold. All other households are classified as 'lower income level'. A finer partition of the income continuum could give even stronger results.

[6]In this paper, we do not focus specifically on gender differences in socioeconomic gradient of obesity. It is a robust finding that women exhibit a steeper socioeconomic gradient, whereas the relationship for men tends to be less steep and curvilinear, although the overall effect is in the same direction. A simple explanation for the curvilinear relationship for men is that the lowest ranked occupations still require a significant amount of physical labour on behalf of men, which could explain why their obesiy levels drop slightly between the fourth and the fifth income quitiles. The steepness of the women's gradient is sometimes explained by reference to social norms, however, this is a question for a future model.

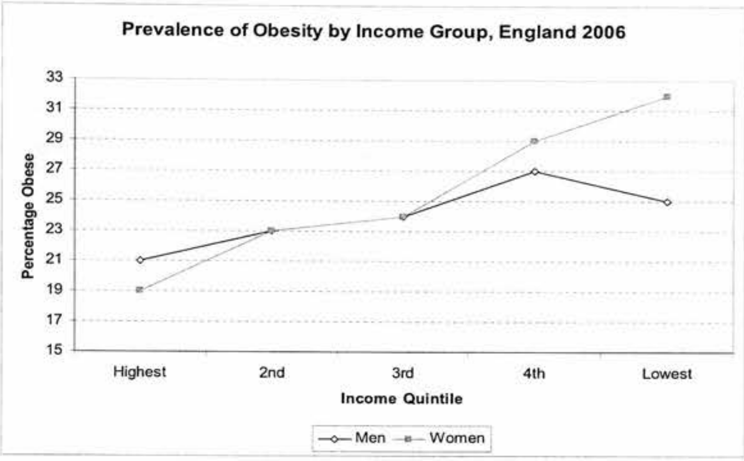


Figure 1b.Obesity over-represneted in lower income groups, especially for women.

What, then, is at the heart of this rise in obesity and why is it overrepresented among the poor? Obesity is defined as an excess of body adiposity. It is measured by the Body Mass Index (weight divided by height squared, kg/m^2) with BMI in the range of 25-30 being overweight, and BMI>30 classified as obesity. The basic laws of thermodynamics dictate that in order to gain weight an individual must experience a positive imbalance between energy intake and energy outlet. In principle, this is a simple equation, but the precise reasons why the imbalance changed dramatically in the 1980s are not fully understood. The analysis is complicated by the fact that the recent rise in obesity could be explained by as little as an average net increase of 100-150 calories a day (Hill et al., 2003).

Data on dietary intake in most developed countries tend to show an upward trend. In the US, dietary surveys and food disappearance data are consistent in indicating an increase in the caloric intake of 200kcal per day over the last 20 years (Popkin et al, 2002; Caballero, 2007). Data for Britain show a roughly 250 calorie rise from 1980 to 2003 (WHO, Europe 2005). The increase in calorie consumption has been attributed to increased consumption of grains, added fats and added sugars (Putnum et al., 1999); to increased consumption of carbohydrates (Finkelstein et al, 2005); to increased consumption of sweetened but energy-dilute beverages (DiMeglio et al., 2000), to increased consumption of energy dense foods (Drewnowski, 2004); to snacking between meals (Cutler et al, 2003); to more frequent visits to and increased portions in fast-food restaurants (French et al., 2000, Nielsen et al, 2003), to increased portions in full-service restaurants (Chou et al, 2004); also to increased portions at home (Nielsen et al, 2003); and even to reduction in smoking (Chou et al, 2004). In other words, although all of these factors and more may contribute to the rise of obesity, the overall picture is far from clear[7].

[7]Stephen J. Dubner even suggests improvements in the plumbing and lavatory systems as possible

On the energy output side, the sedentary lifestyle of the US population as well as many other developed nations, driven by urbanization, changes in transportation systems and the effects of technological change reducing energy expenditure in the workplace, no doubt contributes to the energy imbalance. However, most of these changes pre-date the 1980s and in the US, the sedentary lifestyle was already a concern in the 1950s when the Council on Fitness and Health was created (Caballero, 2007). Cutler, Glaeser and Shapiro (2003) also report that energy expenditure fell substantially before 1975, but has remained roughly constant thereafter. Thus, the precise effects of changes in energy expenditure are also hard to quantify.

What does come out clearly from the vast volume of empirical research on obesity is the fact that food has become abundant, virtually universally available and cheap, with average food prices rising slower than inflation in both the US and Europe. This is particularly true of the energy dense and pre-prepared foods as advances in food processing and storage technology in the 1980s have made sure that these foods have fallen dramatically in relative price. There is also evidence that diets vary by socioeconomic class. Lower income groups are believed to consume more grains, added sugars and fats, and low cost meat (James et al, 1997). Higher income groups tend to consume more fruit and vegetables (De Irala-Estevez et al., 2000). However, given the price differential between fruit and fats, the typical differences by income tend to be relatively small. For example, a European study (De Irala-Estevez) found that the difference in vegetable consumption between the highest and the lowest SES was 17g/person/day for men and 13g/person/day for women. The fruit figures tend to be a little higher, but fruit is also much less representative of diet as a whole. In addition, a study by nearly the same people as the De Irala-Estevez (2000) finds that higher socioeconomic status is also associated with higher consumption of cheese! (Sanches-Villegas, 2003).

These kind of findings underlie the reasoning behind the most important competing explanation for the inverse socioeconomic gradient of obesity. The Drewnowski et al. (2004) hypothesis is that in view of relative prices, the poor income groups cannot afford a lean diet and instead consume the considerably cheaper alternative of energy dense foods, which tends to be strongly associated with positive energy imbalance. Drewnowski shows that energy density and energy cost of food are inversely linked and argues that "the selection of energy dense foods by the food-insecure or low-income consumers may represent a deliberate strategy to save money". We believe that whilst the availability of the low-cost energy dense alternative plays an important role, relative prices alone do not tell the whole story.

First of all, incomes have risen faster than food costs over the last three decades, with expenditure on food as proportion of income declining not only on average but

contributing factor to obesity in "The Strangest Factor Yet for Rising Obesity?", NYT, September 17th 2008.

also across income groups (Offer 2001, Bureau of Labour Statistics). In the UK, the Food Expenditure Survey indicates that whereas in 1957 food and non-alcoholic drinks accounted for the highest proportion of weekly expenditure at 33%, by 2007 this figure was reduced to 15%. Housing expenditure rose from 9% to 19% but did not take up the full slack from reduced food expenditure. The greatest proportional rise has in fact been seen in leisure services, which rose from 6% in 1978 to 15% in 2006. This suggest that the food constraint cannot be more binding today than 30 years ago.

Second, rich in added fats and sugars, energy dense food is not only cheap, it is also highly palatable. It is a standard finding that foods that are energy-dense provide more sensory enjoyment and pleasure than other types of foods (Drewnowski, 1997, 1999; Mela, 1999). Clinical studies also suggest that the most likely targets of food cravings are those foods that contain fat, sugar or both (Yanovski, 2003). This leads us to the more general argument: the abundance of energy-rich palatable food is an incredibly recent phenomenon in the scale of humanity's nearly 200,000 year history, most of which, or at least what we know of it, had been characterised by food scarcity. At times of such scarcity, a preference for energy dense food may have represented an evolutionary advantage (Friedman 1992, Drewnowski, 1995). Fat storing or 'overconsumption' when food was available could have been viewed as a strategy against the times of food insecurity. However, in the age of dietary abundance, such 'thrifty genotype' preference can lead to overeating and, ultimately, obesity (Blundell et al, 1996). For most people, with 'normal' metabolic rates and typical rates of physical activity, not eating oneself into weight gain today requires effort - it requires self-control.

In our view, the problem of self-control lies at the heart of the obesity issue. Indeed, an informed individual with no self-control problem would have no trouble staying thin - he could simply eat less, just enough to maintain the energy balance. However, anyone who has ever tried to lose weight knows well that there is nothing simple about eating less. The estimated $40-$100 billion revenues of the current American dieting industry are but one testament to the difficulty of maintaining a desirable weight (Cutler et al., 2003). Additionally, as well as affecting the poor and the racial minorities, the increase in obesity falls heavily on the right tail of the weight distribution, with already obese people getting disproportionately more obese (Graham and Felton, 2005). This observation offers further support to the self-control hypothesis as people on the right tail of the BMI distribution are already likely to have self-control problems, which would only be exacerbated by the recent developments in the supply of palatable energy-dense foods.

We thus adopt the view that the developments in technology that made food abundant, available and cheap, have also made obesity a problem of self-control. We then develop a model which can explain why low income households can end up with lower self-control in equilibrium. The underlying intuition is that the payoff to over-eating is more or less constant across individuals and independent of wealth, but the overall well-being is increasing in income. Thus, the relative benefit of over-eating compared to that of 'healthy' lifestyle is lower for the higher income groups, making over-eating relatively less attractive.

This is not to say that the poor are to blame for the obesity problem. We do not argue that they make 'bad' choices or have 'wrong' preferences. We argue that their response to the new food environment is rational, as is their decision to accumulate less willpower. However, we also do not argue that this is First Best. A welfare improvement could still be achieved by changing the food environment and reducing the cost of self-control.

3 The Model

In this model, we consider an infinitely-lived agent comprised of two selves: a rational, forward looking self and a myopic alter ego. At any point in time, where time is infinite and continuous, one of the selves is chosen to act; the rational self is selected with probability $p(X(t))$, where $X(t)$ is his stock of willpower. If chosen, the rational self decides how much self-control activity, $I(t)$, to perform. Self-control is effortful and therefore carries a cost, $e(I(t))$. However, by engaging in self-control the rational self can replenish his stock of willpower, which, if left unused, depreciates at rate δ. This has two effects: first, the utility of the rational self, $u(X(t), y)$, is itself an increasing function of willpower; second, having a higher stock of willpower increases the probability that the rational self is chosen to act in the first place, $p(X(t))$ is increasing in $X(t)$. Since one of the main objectives of this paper is to investigate the relationship between income and self-control, we also include income, y, in the utility function; in this chapter, income enters as a parameter. Thus, the net instantaneous utility of the rational self is $u(X(t), y) - e(I(t))$. If the myopic self is chosen, which happens with probability $1 - p(X(t))$, he foregoes any attempt at self-control and instead engages in temptation activity, which yields instantaneous utility, $u(m)$, where m is the temptation payoff to the myopic self, assumed to be constant.[8]

This formulation is somewhat abstract in the sense that we do not specify exactly what constitutes a self-control activity. We argue that everyday life can be viewed as a sequence of self-control decisions: from getting up early to exercise before work, through foregoing the office snacks to not leaving the last piece of work till the morning

[8]These assumptions put a particular structure on our model. The rational and myopic selves essentially have different choice sets. This structure adds to mathematical simplicity of our approach but is not overly restrictive. Although the rational self cannot choose to engage in temptation directly, he can choose to reduce the stock of willpower and thus increase the probability the myopic self will choose temptation for him in the future.

after, self-control is required for a majority of everyday decisions. Thus, the variable $I(t)$ in our model captures how much self-restraint the rational part of the agent chooses to exert on any given occasion. The collection of such self-control decisions can be seen as investment into willpower. We argue further that willpower has a direct effect on agent's (instantaneous) utility. A higher amount of willpower allows the individual to carry out his plans and achieve his goals, undertake and complete long-term projects from acquiring education, to reaching a healthy compromise in marriage or holding on to a job. All of these translate into greater wellbeing (see numerous works by Oswald et al); in fact, much of the economics of happiness shows that variables such as marriage and employment have the most significant effect on happiness; their effect is also considerably greater in magnitude than that of income (Blanchflower and Oswald, 2002). We capture the effect of the major non-monetary, self-control related variables on happiness by introducing willpower, $X(t)$, into the utility function, $u(X(t), y)$[9].

We make the following assumptions about the payoff and cost functions:

Assumption 1

1. *The instantaneous payoff function* $u(X, y)$ *is concave in* X , *with* $\lim_{X\to 0} \dfrac{\partial u(X,\cdot)}{\partial X} = \infty$ *and* $\lim_{X\to\infty} \dfrac{\partial u(X,\cdot)}{\partial X} = 0$
2. *The probability function* $p(X)$ *is concave in* X, *with* $p(0) = 0$ *and* $\lim_{X\to\infty} p(X) = 1$.
3. *The cost function* $e(I)$ *is convex in* I, *bounded and* $e'(I)$ *is bounded above.*

   Moreover, $u(X, y)$, $p(X)$ *and* $e(I)$ *are all* $C^{(2)}$.

The Inada conditions in the first part of this assumption help us get rid of the unnecessary Lagrange multipliers. The second part simply ensures that the probability lies between zero and one. The third part captures what is perhaps the most documented psychological insight about self-control: in the short run, exercising self-control is effortful and becomes progressively more difficult the more self-control is required. We represent this observation with a convex cost of self-control. Notice that this is not exactly the same as Baumeister's argument that repeated self-control tasks weaken one's self-control 'ability' in the short run; nevertheless, our intuition is similar, although we prefer the interpretation of increasing cost of self-control effort. The second psychological insight we incorporate is that in the long run, repeated exercise of self-control can strengthen the individual's willpower reserves, which in our model leads to

[9] As a robustness check, we can respecify the model so that utility is a function of income alone, $u(y)$, and willpower enters into the cost of self-control, $e(I, X)$, with $e_X(I, X) < 0$ and $e_{IX}(I, X) < 0$. In this case, greater willpower reduces the cost of self-control, which is still in line with Baumeister intuition. Under this specification, qualitative results of our model remain unchanged.

an increase in $p(X(t))$, which is an increase in the proportion of time the rational self is 'in control'.

The rational self is sophisticated insofar as he acts like a standard rational expected utility maximiser, taking into account that, with probability $1 - p(X(t))$, the myopic self will be making decisions at any point t in the future. The rational self discounts the future exponentially at rate r. Thus, the objective of the rational self is to choose the optimal rate of self-control to maximise his discounted lifetime utility subject to the equation of motion for willpower:

$$\max_{I} \int_0^{\infty} \left[p(X(t)) \left[u(X(t), y) - e(I(t))\right] + (1 - p(X(t))) \, u(m)\right] e^{-rt} dt \tag{3.1}$$

$$\text{s.t. } \dot{X}(t) = p(X(t))I(t) - \delta X(t), \quad X(0) = X_0, \; X(t) \geq 0$$

It is worth drawing attention to how the formulation of this objective function was arrived at. Suppose time were infinite but discrete; then in every period t, one of the selves would be called upon to make the decision with probability $p(X(t))$. As we take the continuous limit, we assume that the frequency with which decisions are being made is increasing at the same rate with which the time intervals are becoming shorter. In the limit, it's as if the agent were making decisions infinitely frequently. We can therefore write down lifetime utility as the average between the rational and the myopic selves' payoffs, weighted by the respective probabilities of each selves taking the decision. This would be the standard modelling approach in macroeconomics, although one could argue that time intervals could decrease at one rate and decision times could arrive at a different rate, following some stochastic process, for example. The obvious advantage of our approach is that it adds greatly to tractability, and we do not believe that complicating the decisions rate would qualitatively alter our predictions.

The willpower constraint in (3.1), reflects the idea that only the rational self will exercise self-control and replenish the stock of willpower, but the depreciation of willpower occurs regardless of which self is in control. Since we treat willpower as a stock, its value should not be negative. It is easy to see that the $X(t) \geq 0$ constraint is not binding because it can never be optimal for the agent to choose $X(t) = 0$ for any t as our assumptions on the payoff function, namely $\lim_{X \to 0} \dfrac{\partial u(X, \cdot)}{\partial X} = \infty$, ensure that if $X = 0$, an infinitesimally small increase in X would lead to an infinitely large increase in marginal payoff. We do not restrict the value of $I(t)$ to the positive domain so the agent is permitted to disinvest in willpower. Thus, there will be no Lagrange multiplier associated with $I(t)$ and the multiplier for $X(t)$ will be zero.

In its simplified form, Problem 3.1 can we re-written as follows[10]:

[10]We have dropped the time subscribts since this is an infinite horizon optimal control problem in

$$V(X)=\max_{I}\int_0^{\infty}[p(X)(u(X,y)-e(I)-u(m))+u(m)]e^{-rt}dt \tag{3.2}$$

$$\text{s.t. } \dot{X}=p(X)I-\delta X, \quad X(0)=X_0$$

Consider the expression $u(X,y)-e(I)-u(m)$, which is in the objective functional of (3.2). This expression is the difference between the net payoffs to self-control, $u(X,y)-e(I)$, and to temptation, $u(m)$. The economically interesting case is the one in which the long term benefits to self-control are greater than those to temptation. Otherwise, the rational self could run his willpower reserves down to zero, let the myopic self be always in control and always engage in temptation. This is not impossible, but neither is it interesting or plausible. Thus, we assume that for any 'reasonable' cost of self-control, the steady state payoff to self-control exceeds the payoff to temptation.

Assumption 2 *In steady state,*

1. $u(X^S,y)-e(I^S)>u(m)$,
2. $\delta-I^Sp'(X^S)>0$.

The second part of this assumption states that the slope of the $\dot{X}$ isocline at steady state, $\frac{\partial I}{\partial X}=\frac{\delta-p'(X^S)I^S}{p(X^S)}$, should be positive. This means that self-control is increasing with willpower along the $\dot{X}=0$ path, so that maintaining a steady state with higher willpower requires more self-control. In principle, it is possible to have the opposite sign, so that higher willpower leads to less self-control in steady state. This could arise if $p(X)$ was rising sufficiently quickly to compensate for the fall in I. We think the case where the rational self gains control at such at rate that he needs to exert less and less self-control effort as he accumulates willpower is not particularly plausible, so we assume it away.

Proposition 1.1 *Under Assumption* 1, *the solution of the rational agent's optimisation problem approaches a steady state. Under additional Assumption* 2, *this steady state is an asymptotically unstable saddle point if,*

- $\frac{p'(X^S)}{p(X^S)}<\frac{-u_{xx}(X^S,y)}{2u_x(X^S,y)}$, *and only if*
- $2p'(X^S)u_x(X^S,y)<\frac{(r+\delta-p'(X^S)I^S)(\delta-p'(X^S)I^S)e''(I^S)}{p(X^S)}-$
  $((u(X^S,y)-e(I^S)-u(m)+e'(I^S)I^S)p''(X^S)+p(X^S)u_{xx}(X^S,y))$

which time does not enter as an independent variable.

Proof. The proof is a straightforward application of the necessary and sufficient conditions of the infinite time optimal control theory.

Step 1. Write down the current value Hamiltonian with the necessary conditions, and confirm existence:

$$H(X, I, \lambda) = p(X)\left[u(X, y) - e(I) - u(m)\right] + u(m) + \lambda(p(X)I - \delta X) \tag{3.3}$$

$$\frac{\partial H}{\partial I} = -p(X)\left(e'(I) - \lambda\right) = 0 \tag{3.4}$$

$$\dot{\lambda} = r\lambda - \frac{\partial H}{\partial X} = (r+\delta-p'(X)I)\lambda - p'(X)\left[u(X, y) - e(I) - u(m)\right] - p(X)u_X(X, y) \tag{3.5}$$

$$\dot{X} = p(X)I - \delta X \tag{3.6}$$

$$\lim_{t\to\infty} e^{-rt}H(X, I, \lambda) = 0 \tag{3.7}$$

First, we check that the necessary transversality condition (3.7) is satisfied. Define a vector of model parameters and initial conditions $\phi = (\rho, X_0) = (\delta, r, m, y, X_0)$. Assume there exists a solution of the necessary conditions (3.4)-(3.6), which we denote by $(X^*(t,\phi), I^*(t,\phi))$, and a corresponding current value costate variable $\lambda(t,\phi)$, with the property that $(X^*(t,\phi), I^*(t,\phi)) \to (X^s(\rho), I^s(\rho))$ as $t \to \infty$, where $(X^s(\rho), I^s(\rho))$ is the simple steady state solution of the same necessary conditions. We can see from equation (3.4) that $\lambda = e'(I)$; since by Assumption 1, the RHS is bounded and continuous at $(X^s(\rho), I^s(\rho))$, we can say that $\lambda \to \lambda^s$ as $t \to \infty$. Then, consider the infinite time limit of the Hamiltonian:

$$\begin{aligned}\lim_{t\to\infty} H(X^*, I^*, \lambda^*) &= \lim_{t\to\infty}\left(p(X^*)\left[u(X^*, y) - e(I^*) - u(m)\right] + u(m) + \lambda^*(p(X^*)I^* - \delta X^*)\right)\\ &= p(X^S)\left[u(X^S, y) - e(I^S) - u(m)\right] + u(m) + \lambda^S(p(X^S)I^S - \delta X^S) = const.\end{aligned}$$

Since $\lim_{t\to\infty} H(X^*, I^*, \lambda^*)$ exists, we can confirm that

$$\lim_{t\to\infty} e^{-rt}H(X^*, I^*, \lambda^*) = \lim_{t\to\infty} e^{-rt} \lim_{t\to\infty} H(X^*, I^*, \lambda^*) = 0$$

Next, we need to confirm that Mangasarian sufficient conditions are satisfied and so the solution to the necessary conditions is in fact the solution to Problem (3.2). The sufficient conditions in question are that the Hamiltonian, $H(X, I, \lambda)$ is concave along the $(X^*(t,\phi), I^*(t,\phi))$ path[11], and that $\lim_{t\to\infty} e^{-rt}[\lambda(t,\phi)(X^*(t,\phi)-X(t))] \leq 0$. Since we have assumed that as $t \to \infty$, $X^*(t,\phi) \to X^S(\rho)$, $\lambda(t,\phi) \to \lambda^S(\rho)$, $\beta(t,\phi) \to \beta^S(\rho)$ and because all admissible paths of $X(t)$ are bounded, the second sufficient condition is satisfied with equality. To check for concavity of the Hamiltonian, we need to confirm that the Hessian matrix for the Hamiltonian is a negative semi definite, i.e. its eigenvalues are non-positive. With $\lambda(t) = e'(I(t)) \forall t \in [0,\infty)$, from (3.4), the eigenvalues can be written as:

$$\begin{aligned} \nu_1 &= -p(X)e''(I) \qquad (3.8) \\ \nu_2 &= (u(X^S,y) - e(I^S) - u(m) + e'(I^S)I^S)p''(X^S) + 2p'(X^S)u_x(X^S,y) + p(X^S)u_{xx}(X^S,y) \end{aligned}$$

It is clear from (3.8) that $\nu_1 = -p(X)e''(I) < 0$. A sufficient condition for $\nu_2 < 0$ is that $\frac{p'(X^S)}{p(X^S)} < \frac{-u_{xx}(X^S,y)}{2u_x(X^S,y)}$. As we shall see, this is the same sufficient condition that guarantees that the determinant of the Jacobian of the steady state system is negative. If this sufficient condition holds, then the Hamiltonian is in fact strictly concave and the solution is unique.

Step 2. Solve the necessary conditions to find a system of differential equations in (X, I).

After differentiating the first order condition (3.4) with respect to time and substituting into (3.5) we get the following system:

$$\dot{X} = p(X)I - \delta X \qquad (3.9)$$

$$\dot{I} = \frac{-p'(X)(u(X,y) - e(I) - u(m)) + e'(I)(\delta + r - p'(X)I) - p(X)u_x(X,y)}{e''(I)} \qquad (3.10)$$

The steady state occurs when $\dot{X} = 0$ and $\dot{I} = 0$:

$$p(X^s)I^s - \delta X^s = 0 \qquad (3.11)$$

$$-p'(X^S)(u(X^S,y) - e(I^S) - u(m)) + e'(I^S)(\delta + r - p'(X^S)I^S) - p(X^S)u_x(X^S,y) = 0 \qquad (3.12)$$

Step 3. Find the Jacobian of the system (3.9)-(3.10) and ascertain the local stability of the dynamic system evaluated at steady state.

[11]Technically, it is the Lagrangian that should be concave, but since the only Lagrange multiplier in our problem is zero, the condition reduces to the concavity of the Hamiltonian.

$$
\begin{aligned}
J(X^s, I^s) &= \begin{bmatrix} \frac{\partial \dot{X}}{\partial X} & \frac{\partial \dot{X}}{\partial I} \\ \frac{\partial \dot{I}}{\partial X} & \frac{\partial \dot{I}}{\partial I} \end{bmatrix}_{(\dot{X}=0,\ \dot{I}=0)} \\
&= \begin{bmatrix} -\delta + p'(X^S)I^S & p(X^S) \\ -\frac{1}{e''(I^S)}[(u(X^S,y) - e(I^S) - u(m) + e'(I^S)I^S)p''(X^S) & \\ \quad +2p'(X^S)u_x(X^S,y) + p(X^S)u_{xx}(X^S,y)] & r + \delta - p'(X^S)I^S \end{bmatrix}
\end{aligned} \tag{3.13}
$$

Looking at the trace of the Jacobian, $Tr(J(X^s, I^s)) = r > 0$, we can immediately conclude that at least one of the eigenvalues is positive, therefore ruling out asymptotic local stability of the steady state. However, if we can show that the determinant of the Jacobian is negative, that would necessarily mean that the second eigenvalue is negative and so there exists one trajectory which approaches the steady state.

$$
\begin{aligned}
Det[J(X^s, I^s)] &= -(r + \delta - p'(X^S)I^S)(\delta - p'(X^S)I^S) + \\
&\quad \frac{p(X^S)[(u(X^S,y)-e(I^S)-u(m)+e'(I^S)I^S)p''(X^S)+2p'(X^S)u_x(X^S,y)+p(X^S)u_{xx}(X^S,y)]}{e''(I^S)}
\end{aligned} \tag{3.14}
$$

To ascertain the sign of the determinant, let's first have a closer look at the steady state condition (3.12). By Assumption 2, the second part of the first term $(u(X^S, y) - e(I^S) - m) > 0$. In addition, all of $p(X^S)$, $p'(X^S), u_x(X^S, y) > 0$. Then, the first and the third term of expression (3.12) are both negative, meaning that the second term has to be positive, which implies that $\delta + r - p'(X^S)I^S > 0$. Thus, using Assumption 2, we can establish that the first term in $Det[J(X^s, I^s)]$ is always negative. Recalling that $e''(I) > 0$ (from convexity), whereas $p''(X) < 0$, $u_{xx}(X, \cdot) < 0$ (from concavity), we can see that the determinant is negative whenever $2p'(X^S)u_x(X^S, y) + p(X^S)u_{xx}(X^S, y) < 0$. Rearranging the last expression, we get $\frac{p'(X^S)}{p(X^S)} < \frac{-u_{xx}(X^S,y)}{2u_x(X^S,y)}$. This is a sufficient condition for the steady state to be a locally asymptotically unstable saddle point. It is convenient algebraically but is considerably stronger than what is required for a negative determinant.

The necessary condition is

$$
\begin{aligned}
2p'(X^S)u_x(X^S, y) &< \frac{(r+\delta-p'(X^S)I^S)(\delta-p'(X^S)I^S)e''(I^S)}{p(X^S)} - \\
&((u(X^S, y) - e(I^S) - u(m) + e'(I^S)I^S)p''(X^S) + p(X^S)u_{xx}(X^S, y))
\end{aligned}
$$

which is just a rearrangement of (3.14). ■

The sufficient condition of Proposition 1.1 is neat and carries a simple meaning. It states that the rate of increase of the probability function with respect to willpower is less than half the rate of decrease in marginal utility with respect to willpower. This is in parallel to the second part of Assumption 2, namely that the $\dot{X} = 0$ isocline is upward sloping at steady state, which can also be simplified to $\frac{p'(X^S)}{p(X^S)} < \frac{1}{X^S}$. This last expression states that the rate of increase of probability with willpower is bounded above by the reciprocal of willpower. Thus, both Assumption 2 and the sufficient condition of Proposition 1.1 demand that the probability of the rational self being in control does not rise too quickly with willpower; otherwise the rational self may not have the sufficient incentive to invest into willpower by exercising self-control. This seems reasonable enough and it is easy to find functions that would satisfy these conditions[12]. It is important to remember though that this sufficient condition is considerably stronger than what is necessary to guarantee that the steady state is a saddle point. The necessary condition in Proposition 1.1 is far more relaxed, but it also does not lend itself easily to a natural interpretation. Still, we present both for the sake of interest.

The steady state is depicted in Fig.2. The one stable manifold that leads to the steady state has the following properties. If an individual is endowed with high amount of willpower to start with, the optimal path dictates that the individual should disinvest in self-control to drive down willpower to its steady state level. In this case, the stock of willpower declines monotonically until the steady state level is reached. For example, if an individual has had a very strict parental upbringing, he can afford to 'loosen up' and would still reach the steady state. In fact, if an individual complements the initially high endowment of willpower with excessive self-control efforts, he could spiral off into a 'control freak state', forever increasing both self-control efforts and willpower stock. Such out-of-equilibrium behaviour could potentially represent conditions such as anorexia, etc.

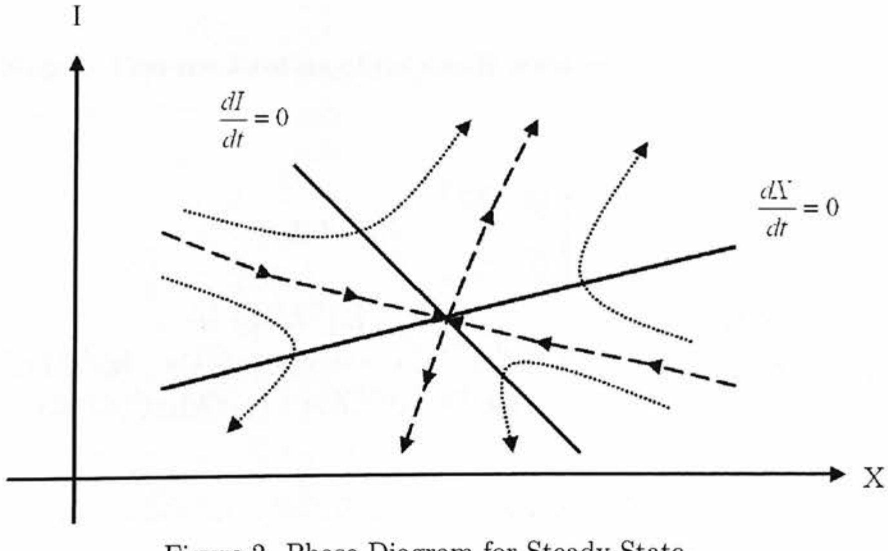


Figure 2. Phase Diagram for Steady State.

If, on the other hand, the initial endowment of willpower is low, the individual should start off investing heavily in self-control, gradually decreasing investment as he

[12] We were able to simulate the model and find the steady state with simple power functions.

approaches the steady state. In this case, the stock of willpower will rise monotonically until the steady state is reached. Thus, our model suggests that an individual who has had a very relaxing childhood, should compensate for his low willpower stock by adopting a strict approach to self-control, straight away. With time, he can afford to relax some of his control rules, but the initial investment must be high. Otherwise, individual runs the risk of ending up in the "no self-control state" with both self-control efforts and willpower tending towards zero.

3.1 Discussion

One of the main goals of this paper is to establish a relationship between income and self-control. We begin by examining the case in which the payoff to temptation activity does not depend on income. This case is most applicable to temptations such as over-eating (or smoking) and thus also to the obesity trends that we aim to explain.

Proposition 1.2 *When the size of temptation payoff is independent of income and with Assumptions 1 and 2, the steady state levels of self-control and willpower are increasing in income if:*

1. *Independence:* $u_{xy}(X, y) = 0$, *or*
2. *Complementarity:* $u_{xy}(X, y) > 0$, *or*
3. *Substitution:* $u_{xy}(X, y) < 0$ *and* $|u_{xy}(X, y)| < \frac{p'(X)}{p(X)} u_y(X, y)$.

Proof. Step 1. Find the Jacobian of the steady state equations (3.11) and (3.12).

$$J(X^s, I^s) = \begin{bmatrix} \frac{\partial \dot{X}}{\partial X} & \frac{\partial \dot{X}}{\partial I} \\ \frac{\partial \dot{I}}{\partial X} & \frac{\partial \dot{I}}{\partial I} \end{bmatrix}$$

$$= \begin{bmatrix} -\delta + p'(X^S)I^S & p(X^S) \\ \begin{array}{c} -[(u(X^S, y) - e(I^S) - u(m) + e'(I^S)I^S)p''(X^S) \\ +2p'(X^S)u_x(X^S, y) + p(X^S)u_{xx}(X^S, y)] \end{array} & (r + \delta - p'(X^S)I^S)e''(I^S) \end{bmatrix} \tag{3.15}$$

This Jacobian is almost exactly the same as $J(X^s, I^s)$ and in fact, $Det[J^S(X^s, I^s)] = e''(I^S)Det[J(X^s, I^s)]$. Therefore, $Det[J^S(X^s, I^s)] < 0$ under the conditions of Proposition 1.

Step 2. To find the effect of a change in income on the steady state, we totally differentiate the steady state equations with respect to income and then solve the following system:

$$\begin{bmatrix} \frac{\partial \dot{X}^S}{\partial X} & \frac{\partial \dot{X}^S}{\partial I} \\ \frac{\partial \dot{I}^S}{\partial X} & \frac{\partial \dot{I}^S}{\partial I} \end{bmatrix} \begin{bmatrix} \frac{\partial X^s}{\partial y} \\ \frac{\partial I^s}{\partial y} \end{bmatrix} = \begin{bmatrix} -\frac{\partial \dot{X}^S}{\partial y} \\ -\frac{\partial \dot{I}^S}{\partial y} \end{bmatrix} \tag{3.16}$$

where the first term is the Jacobian we found in Step 1 and

$$\begin{bmatrix} -\frac{\partial \dot{X}^S}{\partial y} \\ -\frac{\partial \dot{I}^S}{\partial y} \end{bmatrix} = \begin{bmatrix} 0 \\ -(p'(X^S)u_x(X^s, y) + p(X^S)u_{xy}(X^s, y)) \end{bmatrix}.$$

By Cramer's rule,

$$\frac{\partial X^s}{\partial y} = \frac{-p(X^S)(p'(X^S)u_y(X^s, y) + p(X^S)u_{xy}(X^s, y))}{Det[J^S(X^s, I^s)]} \tag{3.17}$$

$$\frac{\partial I^s}{\partial y} = \frac{-(\delta - p'(X^S)I^S)(p'(X^S)u_y(X^s, y) + p(X^S)u_{xy}(X^s, y))}{Det[J^S(X^s, I^s)]} \tag{3.18}$$

We have already established that the denominator is negative for both (3.17) and (3.18). Since $p(X^S) > 0$ and $(\delta - p'(X^S)I^S) > 0$, the sign of both $\frac{\partial X^s}{\partial y}$ and $\frac{\partial I^s}{\partial y}$ is determined by the sign of the expression $(p'(X^S)u_y(X^s, y) + p(X^S)u_{xy}(X^s, y))$. In line with conventional economic theory, we assume that the marginal utility of income, $u_y(X^s, y)$, is positive; it is clear, therefore, that $(p'(X^S)u_y(X^s, y) + p(X^S)u_{xy}(X^s, y)) > 0$ for all X, y as long as $u_{xy}(X^s, y)) \geq 0$ or $u_{xy}(X^s, y)) < 0$ and $|u_{xy}(X, y)| < \frac{p'(X)}{p(X)}u_y(X, y)$. ■

Let's consider each case in turn.

Independence. This is the simplest and perhaps the most important result of this model. All that is needed for steady state self-control and willpower to be increasing with income is that marginal utility of income is positive, which is a rather simple assumption. The intuition for this result comes from the idea that the payoffs to 'temptation' activities such as over-eating or smoking are independent of income but the utility of the rational self is increasing in income. In other words, an individual's enjoyment of the 'sensible' or rational way of life is rising with income. Therefore, for any given level of willpower and self-control, temptation yields a lower relative payoff compared to self-control for people at higher wealth levels, making it relatively less attractive. Notice that this is not the same as saying the rich have a higher payoff to self-control - we have just assumed that marginal benefit of self-control is independent of income. Instead, we are saying that because the poor enjoy lower well-being by nature of lower income, simple temptations such as overeating are more attractive to them.

Complementarity. In this case the cross-partial derivative is positive, $u_{xy}(X^s, y) > 0$, and willpower and wealth reinforce each other so that exerting self-control is more effective at higher levels of wealth and vice versa. For example, it is easy to imagine that increasing an individual's willpower so that he has the patience to study harder and at the same time increasing his income so he can afford a good school would have a positive effect on his welfare. This is in fact the case when marginal utility of willpower is rising with income so that the rich have a greater payoff to self-control.

Substitution. When the cross-partial derivative is negative, $u_{xy}(X^s, y)) < 0$, so that income and willpower act as substitutes, a rise in income would allow an individual to reduce his stock of willpower and still stay at the same level of utility. In this case, if the substitution effect is sufficiently large, an increase in income would lead to a lower self-control steady state. The intuition then is that the poor can compensate for their lack of income by building up the stock of willpower, whereas the rich could substitute their wealth for the self-control effort. More so, one could imagine a utility function for which the cross derivative is negative but its modulus is growing in y. Then, for most income levels, the standard result of higher income leading to higher willpower in steady state would hold. However, after some critical level of y, an increase in income would actually reduce equilibrium willpower and self-control. This could perhaps explain why some people at very high income levels engage in habits incompatible with well-developed willpower.

The question of whether income and willpower act as complements or substitutes is open for debate, and insofar as utility functions might differ across individuals, this could be treated as a source of heterogeneity. The Heckman approach offers strong support for complementarity, but it is also plausible that, particularly at very high wealth levels, the substitution effect might take over. Still, our strongest result so far does not depend on the cross-partial relationship between wealth and willpower, as in the case of independence, a rise in income has an unambiguously positive effect on the steady state.

3.2 Temptation Increasing with Income

Some temptations are increasing with income. It is easy to imagine that the higher income groups, with their superior resources, have access to more ways of 'wasting money'. Expensive shopping trips, state of the art cars or electronics, exotic travel destinations are whims available only to the well-off.

Proposition 1.3 *When the size of the temptation payoff is increasing in income, $m'(y) > 0$, and the sufficient conditions of Proposition 1.1 hold, the steady state levels of self-control and willpower are increasing in income whenever:*

1. *Independence: $u_{xy}(X, y) = 0$ and $u'(m)m'(y) < u_y(X^S, y)$, or*

2. *Complementarity:* $u_{xy}(X,y) > 0$ *and* $u'(m)m'(y) < u_y(X^S,y) + \frac{p(X)}{p'(X)}u_{xy}(X,y)$, *or*

3. *Substitution:* $u_{xy}(X,y) < 0$ *and* $u'(m)m'(y) + \frac{p(X)}{p'(X)}\left|u_{xy}(X,y)\right| < u_y(X,y)$.

Proof is analogous to Proposition 1.2 with one extra derivative and is therefore omitted.

Proposition 1.3 says that the steady state is increasing with income under the conditions of Proposition 1.2 and one extra clause: the temptation utility should not rise too quickly with income. Under independence, if an increase in income raises the utility of the rational self by more than the temptation utility of the myopic self, the steady state levels of self-control and willpower also increase. Thus, an individual now only moves to a higher steady state if the marginal effect of income on the enjoyment of sensible life exceeds the marginal effect on enjoyment of temptation. This may explain why many lottery winners squander their winnings in the first few years and rarely significantly improve their long term well-being. The benchmark condition is somewhat relaxed under complementarity since the extra income augments marginal utility of willpower. With substitution, on the other hand, this benchmark is tightened since the extra income can be substituted for willpower, creating an incentive to lower willpower and self-control.

3.3 Abundance and Obesity

We have argued that advances in the food technology that made rich, energy-dense food abundant, readily-available and often ready-to-eat, have turned obesity into a problem of self-control. It is a common observation in the self-control literature that proximity of temptations or cues associated with it, make such temptation more difficult to resist. In the set-up of our model, it is reasonable to suggest that a dramatic rise in the availability of food has increased the cost of self-control. We augment the $e(I)$ function to $e(I,\alpha)$, where α is the abundance parameter. We assume that abundance creates an upwards shift in the cost of self-control function, $e_\alpha(I^S,\alpha) > 0$, and increases the marginal cost of self-control, $e_{I\alpha}(I^S,\alpha) > 0$.

Proposition 1.4 *Under the sufficient conditions of Proposition 1.1, a rise in α lowers the steady state levels of willpower and self-control.*

Proof. Using the steady state Jacobian of Proposition 2, having replaced $e(I)$ with $e(I,\alpha)$,we solve the following system

$$\begin{bmatrix} \dfrac{\partial \dot{X}^S}{\partial X} & \dfrac{\partial \dot{X}^S}{\partial I} \\ \dfrac{\partial \dot{I}^S}{\partial X} & \dfrac{\partial \dot{I}^S}{\partial I} \end{bmatrix} \begin{bmatrix} \dfrac{\partial X^s}{\partial \alpha} \\ \dfrac{\partial I^s}{\partial \alpha} \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial \dot{X}^S}{\partial \alpha} \\ -\dfrac{\partial \dot{I}^S}{\partial \alpha} \end{bmatrix} \tag{3.19}$$

and find that

$$\frac{\partial X^s}{\partial \alpha} = \frac{p(X^S)(p'(X^S)e_\alpha(I^S,\alpha) + (r+\delta - p'(X^S)I^S)e_{I\alpha}(I^S,\alpha))}{Det[J^S(X^s,I^s)]} \tag{3.20}$$

$$\frac{\partial I^s}{\partial \alpha} = \frac{(\delta - p'(X^S)I^S)(p'(X^S)e_\alpha(I^S,\alpha) + (r+\delta - p'(X^S)I^S)e_{I\alpha}(I^S,\alpha))}{Det[J^S(X^s,I^s)]} \tag{3.21}$$

Recall that $Det[J^S(X^s,I^s)] < 0$. Then since $e_\alpha(I^S,\alpha) > 0$ and $e_{I\alpha}(I^S,\alpha) > 0$, we can conclude that $\frac{\partial X^s}{\partial \alpha} < 0$ and $\frac{\partial I^s}{\partial \alpha} < 0$. ■

Thus, our model predicts that an increase in the availability of food should move an individual to a lower steady state, regardless of wealth level, but for a given level of abundance (α fixed as in previous section), the lower income groups would end up with lower willpower. Notice that although we think it makes sense to assume that both $e_\alpha(I^S,\alpha) > 0$ and $e_{I\alpha}(I^S,\alpha) > 0$, either one of these effects is actually sufficient for willpower to decline with abundance.

4 Evidence from Experimental Psychology

> "When a person trains once, nothing happens. When a person forces himself to do a thing a hundred or a thousand times, then he certainly has developed in more ways than physical. Is it raining? That doesn't matter. Am I tired? That doesn't matter, either. Then willpower will be no problem." Emil Zatopek

The idea that self-control is depleted in the short-run by repeated use finds consistent support in a wealth of experimental evidence. The pattern of 'ego depletion', whereby exerting some self-control effort on the first task weakened the subjects' ability to perform self-regulation on subsequent tasks, has been documented by Baumeister and colleagues in a multitude of experiments, with different manipulations and measures of self-regulation. For example, in Muraven et al. (1998) it was found that the group of people required to engage in thought-suppression exercises in the first part of the experiment gave up on solving (unsolvable) anagrams in the second part of the experiment significantly earlier than the group for which no thought suppression had been required. The exercise in thought suppression also led the subjects to be significantly less able to inhibit their emotional responses to a video clip. In Baumeister et al (1998), it was found that the subjects who had to resist the temptation to eat the chocolate cookies that had been laid out in front of them gave up significantly sooner on the subsequent geometric puzzle task than the participants that had not been subjected to the cookie test. Since then, it has been shown that self-regulation on some initial task weakened the ability to solve reasoning problems (Schmeichel, Vohs, Baumeister, 2003), made people more prone to impulsive purchasing and spending higher amounts in unanticipated buying situations (Vohs and Faber, 2007), increased likelihood of aggressive behaviour (DeWall, Baumeister, Stillman, & Gailliot, 2007), increased likelihood of inappropriate sexual behaviour (Gailliot & Baumeister, 2007). Vohs and Heatherton (2000) also showed that dieters were less able to stick to their diets once they had been exposed to a prior self-control task.

The experiments described above and other similar procedures strongly point to the conclusion that "self-control is impaired in the aftermath of using it" (Baumeister and Vohs, 2007). However, it should be noted that subjects did not simply 'run out' of their self-control stock - rather they were more unwilling to use it. Muraven (1998) found that with sufficiently high incentives subjects were able to exercise self-regulation even after several trials of self-control tasks. One implication of this finding is the support for our modelling of self-control exhaustion as an increase in the cost of the self-control effort. Further, Muraven found support for the idea that people actively 'conserve' their self-control 'energy'. In another experiment, during the second self-regulation task, a group of subjects were told that there was a third task pending. This group showed the biggest decline in self-regulation on the second task. Moreover, their performance on the third task was inversely proportional to how much effort they spent on the second task. In other words, subjects appeared to act strategically with respect to their depleted willpower.

The experimental evidence on the long run properties of willpower has now also come to light, and it is strongly in support of the idea that practising self-control in the short run can strengthen self-control ability, or willpower, in the longer term. First evidence came from an experiment by Muraven, Baumeister and Tice (1999), which required the treatment subjects to spend 2 weeks performing one of three self-control tasks: monitoring and improving posture, regulating mood or monitoring and recording eating. A control group, not required to practice anything, was also included. To identify any changes in the self-control capacity, these students participated in the thought suppression and hand-grip tasks at the beginning and end of the two week period. It was found that the exercising subjects, aside from those engaged in mood regulation, performed significantly better on the hand-grip task at the end of the two weeks than the non-exercising control group. The interpretation of this finding is complicated by the fact that the results were relative and the control group actually got worse at the self-control task. Nevertheless, the study provides preliminary evidence that self-control could be strengthened with exercise.

A number of studies by Oaten and Cheng provide more conclusive evidence. In the 2006 version of the experiment, Oaten and Cheng first measured the subjects' ego depletion between a thought suppression exercise and their subsequent performance on a visual tracking task. Then, they enrolled the subjects on a two months physical exercise programme. After the two months, they re-did the thought suppression and visual tracking tasks and found that ego depletion was significantly reduced. The authors also found that relative to the control group, the exercise group reduced their cigarette, alcohol and junk food consumption. They also studied more and watched less television. The fact that adherence to an exercise regime produced improvements in spheres of life unrelated to physical exercise (such as studying) is of fundamental importance to showing that willpower can grow with exercise over the long run if willpower is viewed as a single resource required for a majority of everyday decisions.

In another experiment, (2007), Oaten and Cheng found that participation in a four months financial monitoring programme not only reduced impulsive purchases and increased the savings rate of the subjects, but also improved their performance on experimental self-regulation tasks, such as visual tracking. Like in the physical exercise regime, participants also reported reduced consumption of caffeine, alcohol and cigarettes, as well as improved eating habits, emotional control and studying efforts. At the same time, measures of perceived stress, emotional distress and self-efficacy remained unchanged, pointing once again to an improvement in underlying willpower. Further evidence on long term effects of repeated exercise of self-control can be found in Oaten and Cheng (2005) and Gailliot, Plant, Butz and Baumeister (2007).

Putting the short-term and the long-term findings together, we believe the view of self-control as resembling a muscle is apt. Repeated exercise becomes increasingly painful in the short run, but a series of such exercises strengthens the self-control reserves, or muscle, in the long run. A sophisticated individual, who is aware of these effects, can act strategically with respect to managing the strength of his self-control reserves.

5 Review of Selected Self-Control Literature

Ever since Shefrin and Thaler (1981, 1988) introduced their planner-doer version of a dual-self agent, the literature on self-control and addiction has seen several interpretations of the dual-self/dual-system initiative. In what is possibly the most well-known dual-self paper, Fudenberg and Levine (2006) formalize the Shefrin-Thaler ideas and develop a model in which the decision is viewed as a game between a sequence of short run myopic selves and a long run rational self. In this two-stage game, first the long run player chooses the self-control action, which can influence the utility function of the myopic self; then the short run player makes the final decision. In Fudenberg and Levine self-control is always possible: at some reduction of utility, the long run self can always implicitly control the choices of the short run self. As a result, equilibrium of the game effectively corresponds to a solution of a single-agent optimisation problem. This model provides an interesting explanation to a number of empirical findings. In particular, Fudenberg and Levine predict that when the short run self has access to wealth, the savings rate is reduced to keep self-control costs low; they also find that the propensity to spend out of unanticipated cash receipts is greater than out of unanticipated bank account receipts and claim that sufficiently small gambles are evaluated with the preferences (over consumption) of the short run self, which are more risk-averse than preferences over long term consumption, therefore explaining Rabin's (2000) paradox of risk aversion in the large and small. Fudenberg and Levine also make predictions in stationary stopping time problems, can incorporate the effect of cognitive load on self-control, and the reduced version of their model is consistent with Gul and Pesendorfer (2001, 2004) axioms. However, some of their main results, particularly in the banking version of the model, depend of the specific timing of when one short-run self is replaced by the next. Whereas Fudenberg and Levine point out that the "attention span" of the short run self would be an interesting subject of investigation, they do not specify a mechanism which would explain the timing of the succession of the short run selves. Although they make some appeal to the idea that large stakes invoke the long run self, it is still unclear why the short runs self cannot simply grab the money and run out of the bank! Interestingly, because the long run self can effectively ration the short run self, the deterministic version of the banking model has an equilibrium equivalent to a model without a self-control problem. Similarly, with no uncertainty, Fudenberg and Levine cannot predict neither Rabin's paradox nor the procrastination results of O'Donoghue and Rabin (1999, 2001).

Benhabib and Bisin (2005) develop another two-system model based on neuroscientific foundations. They assume that agents have the ability to invoke either automatic processes (susceptible to temptation) or controlled processes, which induce the agent to implement a set of goals, determined independently of the specific choice problem. The final decision is governed by a supervisory function which can override the initial response of the automatic process and implement the controlled goal whenever the consequences of the automatic decision become too costly. In the dynamic consumption-saving setting, agents trade off excessive immediate consumption with a consumption-saving rule, requiring an exercise of self-control. The present bias in the model derives from stochastic shocks that enter directly into the utility function, and hence affect the consumption-saving decision. Benhabib and Bisin show that an environment with larger temptations is characterised by a higher probability that self-control is exercised, but on the other hand, agents in such environments set less ambitious goals in the first place.

Similarly, in Loewenstein and O'Donoghue (2007), behaviour is an outcome of the interaction of deliberative processes, which assess choices with a broad goal-based perspective, and affective processes, which incorporate emotions and motivational drives. Loewenstein and O'Donoghue justify the dual-process approach on the basis of the broad neurophysiological division of the brain into the prefrontal cortex and the evolutionary older brain structures. The latter had evolved to promote survival and reproduction and their functions have changed little over time. The prefrontal cortex, on the other hand, captures what seems to be a uniquely human ability to deliberate on the broader consequences of one's actions. It is assumed that the affective system is initially in control but the deliberative system can take over by exerting some effort, or willpower. Unlike most dual-self models, rather than assume that one system is subordinate to the other, Loewenstein and O'Donoghue allow for complex ways in which the two systems interact and influence each other. For example, aside from the actual neural connections between prefrontal cortex and the rest of the brain, there is also considerable evidence that emotional input from the affective system is required for sound deliberative thinking - without it the deliberative system struggles to assess the value of future consequences. They also incorporate Baumeister's intuition that willpower is depleted with use, so that with repeated attempts at self-control the deliberative system has less influence over behaviour. Additionally, the model has scope for considering how the effects of various environmental stimuli, stress and cognitive load affect the relative influence of the two systems. The Loewenstein and O'Donoghue framework is more general than the majority of the literature and can be applied not only to intertemporal choice, but also to risk taking behaviour and social preferences. In intertemporal choice, which is the domain most relevant to our own paper, the model predicts that people who have particularly strong affective reactions to stimuli will exhibit more myopic behaviour, explains why impulsive actions are often associated with strong emotions, and provides a re-interpretation of hyperbolic discounting. This generality comes at a cost though: the model is not fully specified and does not have an explicit solution. Many of the model's implications come directly from the psychological/neurophysiological assumptions made and are not a result of mathematical logic.

In contrast to the above papers, we do not grant the long run self the ability to always control the choices of the myopic self. The fundamental premise of our paper is that people can make systematic mistakes (even with perfect forecast) and the precise purpose of introducing a second self is to capture such 'mistaken' behaviour. We borrow this idea from Bernheim and Rangel (2004) who present a model of addiction in which they argue that firstly, use among addicts is frequently a mistake; secondly, experience with addictive substance sensitizes individual to addictive cues that trigger mistaken usage; and thirdly, addicts understand their susceptibility to cue-triggered mistakes and attempt to manage the process with some degree of sophistication. In their model, upon exposure to environmental cues, individual can sometimes enter a 'hot mode' in which he always consumes the addictive substance irrespective of underlying preferences; the rest of the time individual operates in a 'cold mode' in which he acts like a standard rational agent who takes into account that the 'hot mode' can sometimes occur. Equilibrium behaviour in the model corresponds to the solution of dynamic programming problem (for the 'cold state' self) with stochastic state dependent mistakes (the cue-triggered 'hot modes'). The Bernheim and Rangel paper is complex and intricate and is very specific to addictive behaviour; however, they do provide a tractable model which augments standard economic theory with biological foundations of addiction in a non-trivial way. In equilibrium of their model, particular patterns of consumption depend systematically on the characteristics of the individual, the substance and the environment. For example, Bernheim and Rangel find that when one substance is more addictive than another, then ceteris paribus the more addictive substance is associated with less consumption among relatively new users but with more consumption among the highly experienced users. In contrast to most dual-system models, the authors also make some concrete welfare predictions. Specifically, they find that a beneficial policy intervention potentially exists only in the circumstances when users unsuccessfully attempt to abstain and depend on the usage patterns. For example, they find it optimal to subsidize an addictive substance when the likelihood of use increases with past experience.

We view the dual-self models in general as an alternative to the hyperbolic discounting models in investigating the circumstances in which decisions may diverge from preferences. As discussed in most of the above papers, the existing fMRI evidence from neuroscience suggests that, although still crude, the dual-system approach is a better fit for the modular structure of the brain than the successive multiple selves of hyperbolic discounting. A further advantage of dual-self models is that, in contrast to the multiplicity of infinite time (quasi-) hyperbolic equilibria, these models are usually able to generate a unique prediction. The multiplicity in the hyperbolic model is analogous to the folk-theorem like multiplicity in standard repeated games. In fact, as Krusell, Kuruscu and Smith (2005) point out, this multiplicity is not resolved even by restricting attention to Markov perfect equilibria[13]. In addition, many hyperbolic equilibria impose a peculiar belief structure on the agent which is often counter-intuitive, see for example the cyclical beliefs in the procrastination model of O'Donoghue and Rabin (2001).

Thus, we choose to rely on the dual-self approach which allows us to model systematic mistakes and generates a unique solution. Although the focus of our paper is on the dynamics of willpower and not explicitly on the interaction of various (often external) forces that determine which self or system is in control of decision making,

[13] Yet, whereas in repeated games there are often intuitive reasons for why different sets of strategies can constitute an equilibrium and why several equilibria may be simultaneously reasonable, in the case of one hyperbolic agent, it is not always clear why the different temporal selves cannot renegotiate themselves out of an unattractive equilibrium.

there is a parallel: by exercising self-control in the short run, the rational self can reduce the frequency with which the myopic self makes decisions in the future. However, this effect is specific to our model and is one of the ways in which we differ fundamentally from the above literature.

Ozdenoren, Salant and Silverman (2006) is the only other paper we are aware of that explicitly models the dynamics of willpower depletion. They examine the canonical single-agent fixed horizon cake eating problem with the introduction of a willpower constraint, where the agent has to split his willpower between rationing the cake and some alternative activity. Based on the same psychological evidence and intuitions from Baumeister et al as our own paper, the authors assume that the greater restraint the consumer exercises, the faster his willpower diminishes; and that a given level of consumption depletes willpower faster when the initial reserves of willpower are lower. In addition to incorporating all the standard results of the self-control literature, such as preference reversals and the demand for pre-commitment, this model is also able to explain upward sloping consumption paths. They argue that an agent may increase consumption over time because exercising self-control later, when his stock of willpower is reduced, may require more willpower than exercising the same self-control earlier. In fact, they show that consumption smoothing is optimal only if there is so much willpower to start with that allocating any more towards the intertemporal activity would not increase the utility that can be achieved from it, which is never optimal provided that redirecting more willpower towards some other activity strictly increases utility obtained in that other activity. This is an interesting result and one not shared by other models of self-control, even though the same behaviour could also be explained by models of anticipatory or reference dependent utility. Another interesting result of this paper is the claim that differences in self-control choices of the rich and the poor are a direct reflection of the differences in wealth. In particular, the authors consider two agents who have the same initial willpower, self-control technology and preferences, but differ in the size of the initial cake. They find that the rich and the poor agents will have the same consumption paths in the absolute until the poor agent runs out of cake, which makes the poor agent appear less disciplined as he is consuming at a higher rate. We think that our intuition, which relies on differences in underlying well-being between the rich and the poor, is more plausible.

6 Conclusion

In thinking about the future directions of this research, it might be instructive to consider the experience of some developing countries, such as Mexico, Thailand, Brazil, which have only recently acquired the widespread obesity problem. The emergence of obesity in these countries initially had a greater effect on the higher socioeconomic groups. In fact, Sobel and Stunkard (1989) showed that whereas obesity followed an inverse socioeconomic gradient in the developed world, in the developing countries the association was strongly positive. Recently, however, these trends began to change, with obesity shifting down the socioeconomic ladder. For example, national surveys in Brazil found that while in 1989 obesity in adults was more prevalent in higher socioeconomic status, 10 years later the higher prevalence was observed among the lower socioeconomic groups (Monteiro et al., 2004). In other words, when the new food technologies were first introduced, they were still inaccessibly expensive to the poor, particularly compared to the cheap local produce, and obesity remained the prerogative of the rich. As the food technologies became more ingrained and mass produced food dropped in relative price, obesity became affordable to the poor, and slimness, in turn, became the prerogative of the rich. We think that such trends in obesity over time could be explained by a model of status and social competition, in which self-control has a signalling value. The existing literature has made some headway in that direction, with Avner Offer (2001) being the main proponent of marriage competition driving the changing weight norms across socioeconomic groups. Other works include Graham and Felton (2005) but a complete model is yet to come.

References

[1] Offer, Avner (2001), "Body Weight and Self-Control in the United States and Brtain since the 1950s", Social History of Medicine, vol 14 (1), pp. 79-106

[2] Offer, Avner (2000), "Economic Welfare Measurements and Human Well Being", Discussion Papers in Economic and Social History, No. 34, University of Oxford

[3] Ball, Kylie and David Crawford (2005), "Socioeconomic Status and Weight Change in Adults: A Review", *Social Science and Medicine*, vol 60

[4] Baumeister, R.F., Bratslavsky, E., Muraven, M., & Tice, D.M. (1998), "Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74, 1252-1265.

[5] Baumeister, R. F., Butz, D. A., Gailliot, M. and Plant, E. A. (2007), "Increasing self-regulatory strength via exercise can reduce the depleting effect of suppressing stereotypes," *Personality and Social Psychology Bulletin*, 33, 281-294.

[6] Baumeister, Roy F., DeWall, Nathan C., Stillman, Tyler F., Gailliot, Matthew T. (2007), "Violence Restrained:Effects of Self-Regultation and its Depletion on Aggression," *Journal of Experimental Social Psychology*, Academic Press

[7] Baumeister, R. F. and Gailliot, M. T. (2007), "Self-regulation and sexual restraint: Dispositionally and temporarily poor self-regulatory abilities contribute to failures

at restraining sexual behavior," *Personality and Social Psychology Bulletin*, 33, 173-186

[8] Baumeister, R. F. (1998), "The self," In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), Handbook of social psychology (4th ed.; pp. 680-740). New York: McGraw-Hill.

[9] Baumeister, R. F. (2005), The cultural animal: Human nature, meaning, and social life, New York: Oxford University Press.

[10] Baumeister, R. F., Campbell, J. D., Krueger, J. I. & Vohs, K. D. (2003), "Does high self-esteem cause better performance interpersonal success, happiness, or healthier lifestyles?" *Psychological Science in the Public Interest*, 4, 1-44.

[11] Baumeister, Roy F. and Kathleen D. Vohs (2007), "Self-Regulation, Ego Depletion and Motivation", *Social and Personality Psychology Compass*, vol (1)

[12] Benhabib, Jess and Alberto Bisin (2005), "Modeling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions", *Games and Economic Behaviour*, vol 52, pp. 460-492

[13] Bernheim, Douglas and Antonio Rangel (2004), "Addiction and Cue Triggered Decision Processes", *American Economic Review*, vol 94 (5), pp. 1558 – 1590

[14] Blanchflower, David G. and Andrew J. Oswald (2004), "Well-Being over Time in Britain and the USA", *Journal of Public Economics*, vol 88 (7-8), pp. 1359 – 1386

[15] Blundell JE & King NA (1996), "Overconsumption as a cause of weight gain: behavioural–physiological interactions in the control of food intake (appetite)," In The Origins and Consequences of Obesity (Ciba Foundation Symposium 201), pp. 138–158. Chichester, UK: Wiley.

[16] Blundell JE, Lawton CL, Cotton JR & Macdiarmid JI (1996), "Control of human appetite: implications for the intake of dietary fat," Annual Review of Nutrition 16, 285–319.

[17] Bray, George A. and Claude Bouchard (2008), *Handbook of Obesity: Clinical Applications*, Taylor and Francis Inc

[18] Caballero B (2007), "The Global Epidemic of Obesity: An Overview," *Epidemiology Review*, 29, pp. 1-5

[19] Chou, S.-Y., Grossman, M., & Saffer, H. (2004), "An economic analysis of adult obesity: Results from the behavioral risk factor surveillance system," *Journal of Health Economics*, 23, 565-587.

[20] Cutler, David M., Edward L. Glaeser and Jesse M. Shapiro (2003), "Why have Americans Become More Obese?", *Journal of Economic Perspectives*, vol 17 (3), pp/ 93-118

[21] De Irala-Este´vez J, Groth M, Johansson L, Oltersdorf U, Prattala R & Martınez-Gonzalez MA (2000), "A systematicreview of socio-economic differences in food habits in Europe: consumption of fruit and vegetables," *European Journal of Clinical Nutrition*, 54, 706–714

[22] Department of Health (2006), Health Survey for England

[23] DiMeglio DP & Mattes RD (2000), "Liquid verses solid carbohydrate: effects on food intake and body weight," *International Journal of Obesity and Related Metabolic Disorders*, 24, 794–800.

[24] Drewnowski, Adam (1995), "Energy Intake and Sensory Properties of Food," *American Journal of Clinical Nutrition*, vol 62, pp. 1081-1085

[25] Drewnowski, A. 1997, "Taste preferences and food intake," *Annual Review of Nutrition,* 17: 237-53.

[26] Drewnowski, A. 1999, "Intense sweeteners and energy density of foods: implications for weight control," *European Journal of Clinical Nutrition,* 53: 757-63

[27] Drewnowski, Adam and SE Spencer (2004), "Poverty and Obesity: The Role of Energy Density and Energy Costs", *The American Journal of Clinical Nutrition*, vol 79, pp. 6-16

[28] Finkelstein, Eric A., Christopher J. Ruhm, and Katherine M. Kosa (2005), "Economic Causes and Consequences of Obesity", *Annual Review of Public Health*, vol 26, pp. 239-2572

[29] Flegal, Katherine M., Carroll, Margaret D., Ogden, Cynthia L., and Johnson, Clifford L., (2002), "Prevalence and Trends in Obesity Among US Adults, 1999-2000," *The Journal of the American Medical Association*, 288 (14), 1728-1732

[30] French S. A., Hannan P.J., Jeffery R. W., Murray D. M. and Sherwood NE (2000), "Predictors of weight gain in the Pound of Prevention study" *International Journal of Obesity and Related Metabolic Disorders,* 24, 395–403.

[31] Friedman M., Reed D. and Mela D, (1992), "Sensory and metabolic influences on fat intake," In *Dietary Fats: Determinants of Preference, Selection and Consumption.* Mela D (Ed.). Elsevier Applied Science, NY, USA, 117–137

[32] Fudenberg, Drew and David K. Levine (2006), "A Dual Self Model of Impulse Control", *American Economic Review*, vol 96 (5)

[33] Fogel, Robert W. (1994) "Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy," *NBER Working Papers 4638*, National Bureau of Economic Research

[34] Graham, Carol and Andrew Felton (2005), "Variance in Obesity Across Cohorts and Countries: A Norms-based Explanation using Happiness Surveys", *CSED Working Paper* No. 42

[35] Faruk Gul and Wolfgang Pesendorfer, "Temptation and Self-Control", *Econometrica* 2001.

[36] Faruk Gul and Wolfgang Pesendorfer, "Self control, revealed preference, and consumption choice," *Review of Economic Dynamics*

[37] Hill, James O., Holly R. Wyatt, George W. Reed and John C. Peters (2003), "Obesity and the Environment: Where Do We Go From Here?", *Science*, vol 299

[38] International Obesity Taskforce (2004)

[39] James, W. P., Nelson, M.; Ralph, A. and Leather, S., (1997), "Socio-economic determinants of health," The contribution of inequalities in health. BMJ, 314:1545-1549.

[40] P. Krusell, B. Kuruscu and A. Smith, (2005), "Temptation and Taxation," working paper

[41] Loewenstein, George and Ted O'Donoghue (2007), "The Heat of the Moment: Modeling Interactions Between Affect and Deliberation", *Psychology*

[42] McLaren, Lindsay (2007), "Socioeconomic Status and Obesity", *Epidemiologic Reviews* vol 29, pp. 29 – 48

[43] Mela, D.J. (1999), "Food choice and intake: the human factor," *Proceedings of the Nutrition Society,* 58: 513-21

[44] Monteiro CA, Conde WL, Popkin B. (2004), "The burden of diseasefrom undernutrition and overnutrition in countries undergoingrapid nutrition transition: a view from Brazil," *American Journal of Public Health* 2004; 94(3): 433–4.

[45] Mark Muraven, Dianne M Tice, Roy F Baumeister, (1998), "Self-Control as a limited resource: Regulatory depletion patterns," *Journal of Personality and Social Psychology,* Vol 74 Issue: 3 Pages: 774-789

[46] Muraven, M., Baumeister, R. F., & Tice, D. M. (1999), "Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise," *Journal of Social Psychology,* 139, 446-457

[47] Neel, J.V., A.B. Weder and S. Julius, (1998), "Type II Diabetes, Essential Hypertension, and Obesity as 'Symbols of Impaired Genetic Homeostasis': The 'Thrifty Genotype' Hypothesis Enters the 21st Century", *Perpect Biol Med,* 42, 44-74.

[48] Nielsen, Samara J. and Barry M. Popkin (2003), "Patterns and Trends in Food Portion Sizes, 1977-1998," *Journal of the American Medical Association,* vol 289, pp. 450-453

[49] Oaten, M., & Cheng, K. (2005), "Academic stress impairs self-control," *Journal of Clinical and Social Psychology*, 24, 254-279.

[50] Oaten, M., & Cheng, K. (2006), "Longitudinal gains in self-regulation from regular physical exercise," *British Journal of Health Psychology.*

[51] Oaten, M., & Cheng, K. (2007), "Improvements in self-control from financial monitoring," *Journal of Economic Psychology*, 28, 487-501

[52] O'Donoghue, Ted and Matthew Rabin (1999), "Doing it Now or Later," *The American Economic Review*, vol 89 (1), pp. 103-124

[53] O'Donoghue, T. and M. Rabin (2001). "Choice and Procrastination," *Quarterly Journal of Economics*, 116, 121-160.

[54] Offer, Avner (2001), "Body Weight and Self Control in the United States and Britain since the 1950s", *Social History of Medicine*, vol 14 (1), pp. 79-106

[55] Ozdenoren, Emre, Stephen Salant and Dan Silverman (2006), "Willpower and the Optimal Control of Visceral Urges", *NBER Working Paper 12278*

[56] Popkin, Barry M., Samara J. Nielsen, Siega-Riz A.M. (2002), "Trends in energy intake in the US between 1977 and 1996: similar shifts seen across age groups," *Obesity Research* 10:370-378.

[57] Putnam, Judy . and Shirley Gerrior (1999), "Trends in the U.S. Food Supply, 1970-97," *America's Eating Habits: Changes and Consequences*, Washington: USDA ERS, 1999.

[58] Rabin, Matthew (2000), "Risk Aversion and Expected Utility Theory," *Econometrica*, vol 68 (5)

[59] Sanchez-Villegas A, Delgado-Rodriguez M, Martinez-Gonzalez MA & De Irala-Estevez J (2003), "Gender, age, socio-demographic and lifestyle factors associated with major dietary patterns in the Spanish Project SUN," *Clinical Nutrition*, 57, 285–292.

[60] Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003), "Ego depletion and intelligent performance: Role of the self in logical reasoning and other information processing," *Journal of Personality and Social Psychology*, 85, 33-46

[61] Shefrin, H. M. and Thaler R. H. (1981), "An Economic Theory of Self Control," *Journal of Political Economy*

[62] Shefrin, H. H. and Thaler, R. H. (1988), "The behavioral life-cycle hypothesis" *Economic Inquiry*, 26 , 609-643

[63] Sobal, J. and Stunkard A. J. (1989), "Socioeconomic Status and Obesity: A Review of the Literature," *Psychological Bulletin*, vol 105 (2), 260-275

[64] Swinburn, B. A., Egger, G. J. and Raza, F. (1999), "Dissecting Obesogenic Environments: The Development and Application of a Framework for Identifying and Prioritising Environmental Interventions for Obesity," *Preventative Medicine*, 29, 563–570.

[65] Ulijaszek, S.J. (2007), "Obesity: A Disorder of Convenience", *Obesity Reviews*, 8 (Suppl.1), 183-187.

[66] Vohs, K.D. & Heatherton, T.F. (2000), "Self-regulatory failure: A resource depletion approach," *Psychological Science*, 11, 249-254.

[67] Vohs, Kathleen D. and Ronald J. Faber (2007), "Spent Resources: Self-Regulatory Resource Availability Affects Impulse Buying," *Journal of Consumer Research*, 33 (March), 537-547.

[68] Vohs, Kathleen D. and Todd F. Heatherton (2004), "Ego Threat Elicits Different Social Comparison Processes Among High and Low Self-esteem People: Implications for Interpersonal Perceptions," *Social Cognition*, 22 (February), 168–191.

[69] Wardle, Jane, Jo Waller and Martin J. Darvis (2002), "Sex Differences in the Association of Socioeconomic Status with Obesity", *American Journal of Public Health*, vol 92 (8)

[70] World Health Organization. (2004), "Obesity and overweight," Geneva: WHO

[71] Yanovski, Susan (2003), "Sugar and Fat: Cravings and Aversions," *The Journal of Nutrition*, vol 133, pp. 835-7

Part 2

A Model of Job Autonomy and Self-Control

Abstract

In this chapter, we develop a model of job autonomy, human capital and self-control which aims to explain the effect of different types of occupations on self-control outcomes, which is distinct from the pure income effect of wages. Jobs differ in the degree of autonomy placed on the worker. We argue that successful performance in autonomous jobs requires the kind of human capital, acquiring which demands exercise of self-control in the first place. Accumulating such human capital then has spill-over effects on individual's level of willpower in other areas of his life. We show that an increase in the degree of job autonomy in fact increases the steady state levels of willpower, self-control and human capital. Increasing the return to human capital has a similar effect. We also find an upper bound for marginal cost of self-control for which a small increase in autonomy increases agents' experienced welfare in steady state.

1 Introduction

It is now an established empirical fact that obesity has risen dramatically in the developed world over the last 30 years, with the effects of rising weights becoming increasingly detrimental to health (WHO, 2004; HSE, 2006; Flegal et al, 2002; Bray et al, 1998). It is also a well established fact that the burden of obesity falls disproportionately on the least privileged social groups (Sobal and Stunkard, 1989; Healthy People 2010 Report; HSE, 2006). In our previous chapter, we argued that, developed in times of food insecurity, the human genotype may be maladapted to the environment of food abundance (Ulijaszek, 2007), and therefore advances in food technology that have made food abundant, readily-accessible and cheap have also made obesity a problem of self-control. We then developed a dynamic model of willpower which predicts lower self-control outcomes for the lower income groups and thus goes some way towards explaining the inverse socioeconomic gradient of obesity.

However, income itself is only one component of socioeconomic status, which also typically comprises education and occupation. Already in 1989, in their seminal review of 144 obesity studies, as well as showing that groups of lower socioeconomic status were at higher risk of obesity, especially for women, Sobal and Stunkard indicated that results could vary according to the measure of socioeconomic status being used. Since then, studies that consider the components of economic status separately often find the strongest and most consistent inverse relationship of obesity with education (Wardle et al, 2002; McLaren, 1997) and occupation (Ball and Crawford, 2005). This suggests that the income effect alone does not tell the whole story.

In this chapter, we develop a model of job autonomy, human capital and self-control which aims to explain the effect of different types of occupations on self-control outcomes, which is distinct from the pure income effect of wages. Jobs differ in the skills they require; they also differ in the degree of autonomy and self-determination entrusted to the worker. We argue that jobs may also differ in how the relevant skills can be acquired. Accumulating human capital requires effort, but the amount and type of effort may differ across skills. Highly autonomous jobs, which bestow a lot of independence on the worker, may allow for flexibility, creativity and individualised approach, but will also require the individual to exercise discipline and self-control. To become 'productive' at this kind of job, the individual will need to invest in self-control effort. If in addition, the autonomous job is relatively skilled, acquiring the relevant skills would also require self-control effort from the individual. Being 'productive' in a more routine job, which sets stricter guidelines on how the job is to be performed, is more repetitive, or simply has a high level of monitoring tends to demand less self-control effort on behalf of the individual. It is easy to imagine that more routine jobs entail the type of skills which can be more easily picked up "on the job", through learning by doing or formal training. Accumulating the necessary human capital would then also require less self-control effort. The idea is that if successful performance on a job requires self-control or demands the kind of skills, accumulating which requires exercise of self-control effort in the first place, this effort will have spillover effects, leading the individual to accumulate more willpower and be more 'in control' in all areas of his life[14].

We model autonomy as reinforcing self-control efforts in skill accumulation and show that a rise in the degree of job autonomy raises the levels of willpower and self-control in steady state. In our set-up, a rise in autonomy also raises the amount of human capital the individual chooses to accumulate, as autonomy provides the incentive for extra self-control effort which leads to greater skills. Insofar as greater skills command a higher wage, an increase in job autonomy also raises individual's earnings, and therefore consumption.

As in Chapter 1, our model is based on the premise that people can make systematic

[14]This model can equally be re-interpreted as a model of education. Insofar as education is not only about what one learns, but also the process through which an individual gains knowledge and understanding, this process of obtaining education usually demands self-control. People with high levels of education will have invested a lot of self-control effort into learning, which according to our theory, should spill-over into greater willpower in other areas of their life.

mistakes; reducing the likelihood of such mistakes amounts to self-control in our approach. Thus we model the individual as comprised of two selves: rational and myopic. Decisions of the myopic self represent the propensity for systematic mistakes. The rational self, on the other hand, is forward looking and sophisticated in the sense that he understands the process through which the myopic self might come to make decisions, and acts to maximise lifetime utility, taking the myopic self into account[15]. Since the only function of the myopic self is to represent systematic mistakes, we maintain the standard assumption that individual has a single set of preferences, which allows us to use the agent's experienced utility as the welfare criterion. We present an upper bound on the marginal cost of self-control for which an increase in willpower associated with greater income or autonomy, results in a welfare improvement for the agent.

2 The Autonomous Jobs Model

In this model, we continue to consider an infinitely-lived individual, comprised of two selves: a rational forward looking self and a myopic alter ego. At any point in time, where time is infinite and continuous, one of the selves is chosen to act; the rational self is selected with probability $p(X(t))$, where $X(t)$ is the individual's stock of willpower. As in Chapter 1, willpower grows with self-control effort, $I(t)$, and depreciates at rate δ if left unused. The new feature of this model is that the individual now also works. Work requires skills, $K(t)$, which reflect the agent's productivity, and pays a wage, $w(K(t))$. Skills, or human capital, accumulate as a function of self-control effort, $g(I(t), \theta)$, where $g(\cdot)$ is concave in self-control $I(t)$, and θ is a parameter which captures the degree of worker's independence, flexibility or autonomy. We model autonomy as reinforcing the self-control effort in skills accumulation and so assume that $g_{I\theta}(I(t), \theta) > 0$. Human capital depreciates over time at rate δ (knowledge and skills are forgotten with time if not used). For simplicity, we assume the same depreciation rate for human capital and willpower. Finally, the individual derives utility $u(\cdot)$ from consumption; the rational self chooses his level of consumption, $c(t)$; the myopic self consumes the same constant amount, m.

Then, if chosen, the rational self decides how much self-control activity, $I(t)$, to perform, and how much to consume $c(t)$. The choice of self-control effort now has two effects. Firstly, the exercise of self-control allows the individual to accumulate skills $K(t)$, which lead to a wage $w(K(t))$, which in turn allows for the individual's consumption. Secondly, the exercise of self-control increases the stock of willpower,

[15] We borrow the "systematic mistake" interpretation of the myopic self from Bernheim and Rangel, 2004.

$X(t)$, which increases the probability that the rational self is chosen to act in the first place.

Individual's consumption must satisfy the lifetime budget constraint, so that the discounted sum of lifetime expected consumption between the two selves should equal the discounted sum of lifetime earnings[16]:

$$\int_0^\infty \left[p(X(t))c(t) + (1 - p(X(t)))m\right] e^{-rt} dt = \int_0^\infty w(K(t))e^{-rt} dt \tag{2.1}$$

Since savings have no intrinsic value in this model, in the absence of a myopic self, in equilibrium or steady state of this model, the rational self would simply consume his wage. The myopic self represents mistakes in consumption; we assume he 'overconsumes', so that $m > c^S$, and in fact, $m > w(K^S)$. Thus, if the myopic self is in control, he consumes more than the individual's earnings. The rational self then adjusts his consumption plan below $w(K^s)$ to account for the 'overconsumption' of the myopic self. In some sense, the rational self 'saves' to compensate for the myopic self. The assumption that m is constant is somewhat stylised; it is as if there exists a bundle of temptation goods that the myopic self consumes regardless of his wage. This means that an individual with a lower wage will 'overconsume' a relatively greater amount than someone with a higher wage and consequently will have to compensate by adjusting the rational self's consumption even further downwards. Since this kind of structure is imposed on the decision maker, we do not view 'overconsumption' on its own as a measure of self-control in this model. The degree of willpower and self-control is measured only by the variables X and I. To this extent, it is important that our structure does not bias a relatively poor individual towards lower self-control; if anything, the bias is in the opposite direction. Consider a relatively poor individual with $w(K_L)$: every time the myopic self takes control and consumes m, the rational self has to compensate with a low c_L. With a concave utility function, this is costly for the individual as the utility he loses with every unit of foregone consumption when moving to c_L, which lies on the steep part of his utility function is greater than the utility he gains for every unit of extra consumption at m, which is on the shallower part of utility. A relatively rich individual, with $w(K_H)$, would not have to travel as far down his utility function to compensate for the myopic self, so his marginal cost of overconsumption is lower. Thus, the relatively poor individual should have more incentive to reduce the probability of the myopic self taking control, and the only way to do that is to accumulate more willpower[17].

[16] We write budget constraint as satisfied with equality since in the steady state of this model, the agent has no incentive to save, and should consume all his earnings as long as satiation point is not reached.

[17] Of course, letting consumption of the myopic self increase with earnings, $m(w(k))$, would be a more realistic representation. However, we find that in that case, the main results remain qualitatively very

The two selves share the same utility function, which implies that $u\left(c^S\right) < u\left(m\right)$ since $c^S < m$. This is in contrast to our baseline model, described in Chapter 1, where we assumed that, in steady state, the utility of the self-control state was greater than of temptation state, $u(X^S, y) - e(I^S) > u(m)$. The logic was that, otherwise, the rational self could run his willpower reserves down to zero and mimic the myopic self. In this model, the logic is different. The myopic self achieves greater utility than the rational self, but only the rational self puts the effort in to gain skills that earn him wages. Thus, if the rational self were to mimic the myopic self and run willpower reserves down to zero, his human capital would also depreciate to zero, and he would no longer be earning wages. His consumption would then also have to go to zero. For any 'reasonable' cost of self-control, this cannot be optimal[18].

We make the following assumptions on the structure of payoff and cost functions:

Assumption 1

1. *The payoff function* $u(c)$ *is concave in* c , *and* $\lim_{c\to 0} \dfrac{\partial u(c)}{\partial c} = \infty$ *and* $\lim_{c\to\infty} \dfrac{\partial u(c)}{\partial c} = 0$. *Similarly, for* $u(m)$.
2. *The probability function* $p(X)$ *is concave in* X, *with* $p(0) = 0$ *and* $\lim_{X\to\infty} p(X) = 1$.
3. *The cost function* $e(I)$ *is convex in* I, *bounded and* $e'(I)$ *is bounded above.*
4. *The skills accumulation function* $g(\cdot)$ *is concave in* I, *and bounded above*; $g(0) = 0$, $g_{I\theta}(I,\theta) > 0$, $g_\theta(I,\theta) = \epsilon > 0$.

   Moreover, $u(c)$, $p(X)$, $e(I)$ *and* $g(I,\theta)$ *are all* $C^{(2)}$.

In parallel to Assumption 1 of Chapter 1, we assume that utility is concave in consumption and probability is concave in willpower, and that the Inada conditions are satisfied, which allows us to get rid of the multipliers associated with a potential zero solution, which is not economically interesting. The convexity assumption on the cost of self-control is once again driven by the well-documented psychological evidence that repeated exercise of self-control becomes progressively more difficult in the short run (Baumeister and Vohs, 2007). In addition, we also assume that the skills accumulation function is concave in the self-control effort and that self-control and autonomy act as complements; for now we also assume that autonomy on its own has a positive effect on human capital, but this effect is negligible. This represents the situation in which an increase in autonomy, or the degree of self-determination or flexibility, has a positive

similar to the simpler set up (with some additional conditions, naturally) but the increase in algebraic complication is considerable. Thus, we stick to the simpler version with $m = const$.

[18] Given a constant m, the budget constraint would also no longer be satisfied.

effect on workers' skills; we maintain that this effect is small so as not to bias our results.

The sophisticated rational self, who discounts the future at rate r, maximises the discounted lifetime utility subject to the equation of motion for willpower, the accumulation of human capital equation and the budget constraint, taking into account that, with probability $(1 - p(X(t)))$, the myopic self will be making decisions in the future:

$$\max_{I,c} \int_0^\infty \left[p(X(t))(u(c(t)) - e(I(t))) + (1 - p(X(t)))u(m)\right] e^{-rt} dt \tag{2.2}$$

$$\begin{aligned} s.t.\ \dot{X} &= p(X(t))I(t) - \delta X(t),\ \forall X \in [0, \infty),\quad X(0) = X_0 \\ \dot{K} &= p(X(t))g(I(t), \theta) - \delta K(t),\ \forall K \in [0, \infty),\ K(0) = K_0 \\ \int_0^\infty \left[p(X(t))c(t) + (1 - p(X(t)))m\right] e^{-rt} dt &= \int_0^\infty w(K(t)) e^{-rt} dt \end{aligned}$$

The objective functional in (2.2) is the average of the net utilities of the rational and the myopic selves, weighted by the probabilities of the rational and the myopic selves making the decision. The equations of motion for willpower and human capital reflect the idea that only the rational self exercises self-control and therefore accumulates skills and willpower, but the depreciation of both occurs regardless of which self is in control. The explicit introduction of the budget constraint makes this an isoperimetric problem. To solve this problem, we introduce a new variable

$$Q = -\int_0^t \left[p(X(t))(c(t) - m) + m - w(K(t))\right] e^{-rt} dt \tag{2.3}$$

where the expression under the integral is just a rearrangement of last constraint in (2.2) with all terms on the left hand side. Then $\dot{Q} = -\left[p(X(t))(c(t) - m) + m - w(K(t))\right] e^{-rt}$, with $Q(0) = 0$ and $Q(\infty) = 0$, which follows directly from substituting the integration limits into (2.3).

We strive to retain the generality of this model, but for the sake of tractability, we make one further simplifying assumption. We assume that wage is directly proportional to the amount of human capital.

Assumption 2 *Wage is linear in human capital, $w(K) = wK$.*

In simplified form, the rational self's problem can be written as[19]:

$$V(X,K) = \max_{I,c} \int_0^\infty \left[p(X)(u(c) - e(I) - u(m)) + u(m)\right] e^{-rt} dt \tag{2.4}$$
$$s.t.\ \dot{X} = p(X)I - \delta X, \quad X(0) = X_0$$
$$\dot{K} = p(X)g(I,\theta) - \delta K, \quad K(0) = K_0$$
$$\dot{Q} = -\left[p(X)(c-m) + m - wK\right] e^{-rt}, Q(0) = 0, Q(\infty) = 0$$

Proposition 2.1 *The solution of the rational self's optimisation problem (2.4) approaches a steady state. This steady state is an asymptotically unstable saddle point, with two positive and two negative eigenvalues, if* $\delta - I^S p'(X^S) > 0$.

The proof is straightforward but algebraically involved and can be found in the Appendix to this chapter. As in the baseline model, we assume that the $\dot{X} = 0$ isocline is upward sloping in the (X, I) plane, and therefore $\delta - I^S p'(X^S) > 0$, so that higher willpower steady states require higher self-control to sustain them. Recall that if $p(X)$ is rising sufficiently quickly to compensate for the fall in I, it is possible to have the opposite sign, which in this case would also give a unique path that converges to steady state. However, we continue to think that the positive slope is the more natural interpretation, more so even than uniqueness, and hence proceed with Assumption 3.

Assumption 3 $\delta - I^S p'(X^S) > 0$

Corollary 2.1 *The steady state consumption plan of the rational self is* $c^* = m - \frac{m - wK^S}{p(X^S)}$. *It is increasing in steady state levels of human capital and willpower, and decreasing in the consumption of the myopic self. The path of consumption is constant, i.e. the agent 'jumps' to* c^* *at* $t = 0$, *and uses the self-control variable to adjust to the steady state.*

Proof. Write down the current value Hamiltonian and the necessary conditions:

$$\begin{aligned} H(X,K,I,C,\alpha,\beta,\gamma) &= p(X)(u(c) - e(I) - u(m)) + u(m) + \alpha(p(X)I - \delta X) \\ &\quad \beta(p(X)g(I) - \delta K) - \gamma(p(X)(c-m) + m - w(K)) \end{aligned} \tag{2.5}$$

[19] As before, we drop the time subscripts since this is an infinite time problem in which time does not enter as an independent variable.

$$\frac{\partial H}{\partial c} = p(X)(u'(c) - \gamma) = 0 \tag{2.6}$$

$$\frac{\partial H}{\partial I} = p(X)(\alpha - e'(I) + \beta g_I(I, \theta)) = 0 \tag{2.7}$$

$$\dot{\alpha} = (r + \delta - ip'(X))\alpha - p'(X)(u(c) - u(m) - e(i) - \beta g(I, \theta) + (c - m)\gamma) \tag{2.8}$$

$$\dot{\beta} = (r + \delta)\beta - \gamma w \tag{2.9}$$

$$\dot{\gamma} = -\frac{\partial H}{\partial Q} = 0 \tag{2.10}$$

$$\lim_{t \to \infty} e^{-rt} H(X, K, I, C, Q, \alpha, \beta, \gamma) = 0 \tag{2.11}$$

From condition (2.6), it is obvious that $\gamma = u'(c)$. Differentiating both sides with respect to time, we get

$$\dot{\gamma} = u''(c)\dot{c} \tag{2.12}$$

However, from condition (2.10), we can see that $\dot{\gamma} = 0$, since the Hamiltonian is not a function of Q, implying that $\dot{c} = 0$, since $u''(c) < 0\ \forall c$. Thus, consumption follows a constant path. That the multiplier on the integral constraint is constant over time is a standard feature of isoperimetric problems. Notice also, that although $H(X, K, I, C, \alpha, \beta, \gamma)$ is a current value Hamiltonian, γ is not a current, but rather the present value multiplier since the constraint $\dot{Q} = -\left[p(X)(c - m) + m - wK\right] e^{-rt}$ already contains the discount term.

Next, recall that in steady state, all time paths must be constant, including $\dot{Q}$. Solving this expression for c gives $c^* = m - \frac{m - wK^S}{p(X^S)}$. It is straightforward to see that $\frac{\partial c^*}{\partial K^S} = \frac{w}{p(X^S)} > 0$, $\frac{\partial c^*}{\partial X^S} = \frac{m - wK^S}{p(X^S)^2} > 0$. ■

The fact that consumption turns out to be constant is not all that surprising considering that it does not enter into either of the state equations. Since consumption is not one of the levers through which an individual can adjust his levels of willpower or human capital, there is no reason why the agent would find it optimal to consume at any level other than the steady state level while adjusting to steady state. The consumption of the rational self is increasing in willpower since more willpower means that the individual makes myopic mistakes less frequently, i.e. the opportunity for the myopic self to 'overconsume' is reduced, therefore reducing the need for the rational self to compensate by adjusting his consumption downwards. Hence, with higher levels of willpower, the rational self can enjoy higher levels of consumption whilst he is in control, which also achieves a greater degree of consumption smoothing.

2.1 Comparative Analysis

In this section we examine the effects that changes in autonomy and wages have on the steady state.

Proposition 2.2 *Under Assumptions 1, 2 and 3, an increase in job autonomy θ increases steady state willpower, X^S, self-control, I^S, and human capital, K^S.*

Proof. Step 1. The steady state occurs when the following system of equations holds simultaneously:

$$p(X^S)I^S - \delta X^S = 0 \tag{2.13}$$

$$p(X^S)g(I^S,\theta) - \delta K^S = 0 \tag{2.14}$$

$$(r+\delta)\beta - wu'(c^*) = 0 \tag{2.15}$$

$$\begin{matrix} p'(X^S)(u(c^*) - u(m) - e(I^S) + \beta^S g(I^S,\theta) + (m-c^*)u'(c^*)) - \\ (r+\delta - I^S p'(X^S))(e'(I^S) + \beta^S g_I(I^S,\theta)) \end{matrix} = 0 \tag{2.16}$$

Step 2. Find the Jacobian of the steady state system and ascertain the sign of its determinant.

$$J_S(\dot{X}=0,\ \dot{K}=0, \dot{\beta}=0, \dot{I}=0) = \tag{2.17}$$

$$\begin{bmatrix} -\delta + I^S p'(X^S) & 0 & 0 & p(X^S) \\ p'(X^S)g(I^S,\theta) & -\delta & 0 & p(X^S)g_I(I^S,\theta) \\ 0 & 0 & r+\delta & 0 \\ \begin{matrix} p''(X^S)(u(c^*) - u(m) \\ -e(I^S) + I^S e'(I^S) \\ +\beta^S(g(I^S,\theta) - I^S g_I(I^S,\theta)) \\ +(m-c^*)u'(c^*)) \end{matrix} & 0 & \begin{matrix} g(I^S,\theta)p'(X^S) + \\ (r+\delta - I^S p'(X^S))\times \\ g_I(I^S,\theta) \end{matrix} & \begin{matrix} -(r+\delta - I^S p'(X^S)) \\ (e''(I^S) \\ -\beta^S g_{II}(I^S,\theta)) \end{matrix} \end{bmatrix}$$

The determinant of the steady state Jacobian J_S:

$$Det[J_S] = -\delta(r+\delta)\begin{pmatrix} -p(X^S)p''(X^S)(u(c^*) - u(m) + (m-c^*)u'(c^*) - e(I^S) + I^S e'(I^S) + \\ \beta^S(g(I^S,\theta) - I^S g_I(I^S,\theta)) \\ +(\delta - I^S p'(X^S))(r+\delta - I^S p'(X^S))(e''(I^S) - \beta^S g_{II}(I^S,\theta)) \end{pmatrix} \tag{2.18}$$

Consider the sign of $(u(c^*) - u(m) + (m-c^*)u'(c^*) - e(I^S) + I^S e'(I^S) + \beta^S(g(I^S,\theta) - I^S g_I(I^S,\theta))$. By concavity of $u(\cdot)$, and because $c^* < m$, we can say that $(m-c^*)u'(c^*) > u(m) - u(c^*)$ and so $u(c^*) - u(m) + (m-c^*)u'(c^*) > 0$. Similarly, by concavity of $g(\cdot)$, $g(I^S,\theta) - I^S g_I(I^S,\theta) > 0$ and by convexity of $e(\cdot)$, $e(I^S) - I^S e'(I^S) < 0$. Putting these together, we can see that $(u(c^*) - u(m) + (m-c^*)u'(c^*) - e(I^S) + I^S e'(I^S) + \beta^S(g(I^S,\theta) - I^S g_I(I^S,\theta)) > 0$. The expression $(e''(I^S) - \beta^S g_{II}(I^S,\theta)) > 0$ follows immediately from convexity of $e(\cdot)$ and concavity of $g(\cdot)$. Then, under the assumption of $\delta - I^S p'(X^S) > 0$, we can conclude that $Det[J_S] < 0$.

Step 3. To find the effect of a change in autonomy on the steady state, totally differentiate the steady state equations (2.13)-(2.16) with respect to θ, and solve the following system:

$$\begin{bmatrix} \frac{\partial \dot{X}^S}{\partial X} & \frac{\partial \dot{X}^S}{\partial K} & \frac{\partial \dot{X}^S}{\partial \beta} & \frac{\partial \dot{X}^S}{\partial I} \\ \frac{\partial \dot{K}^S}{\partial X} & \frac{\partial \dot{K}^S}{\partial K} & \frac{\partial \dot{K}^S}{\partial \beta} & \frac{\partial \dot{K}^S}{\partial I} \\ \frac{\partial \dot{\beta}^S}{\partial X} & \frac{\partial \dot{\beta}^S}{\partial K} & \frac{\partial \dot{\beta}^S}{\partial \beta} & \frac{\partial \dot{\beta}^S}{\partial I} \\ \frac{\partial \dot{I}^S}{\partial X} & \frac{\partial \dot{I}^S}{\partial K} & \frac{\partial \dot{I}^S}{\partial \beta} & \frac{\partial \dot{I}^S}{\partial I} \end{bmatrix} \begin{bmatrix} \frac{\partial X^S}{\partial \theta} \\ \frac{\partial K^S}{\partial \theta} \\ \frac{\partial \beta^S}{\partial \theta} \\ \frac{\partial I^S}{\partial \theta} \end{bmatrix} = - \begin{bmatrix} \frac{\partial \dot{X}^S}{\partial \theta} \\ \frac{\partial \dot{K}^S}{\partial \theta} \\ \frac{\partial \dot{\beta}^S}{\partial \theta} \\ \frac{\partial \dot{I}^S}{\partial \theta} \end{bmatrix}$$

where the first term on LHS is just J_S and the RHS is

$$- \begin{bmatrix} \frac{\partial \dot{X}^S}{\partial \theta} \\ \frac{\partial \dot{K}^S}{\partial \theta} \\ \frac{\partial \dot{\beta}^S}{\partial \theta} \\ \frac{\partial \dot{I}^S}{\partial \theta} \end{bmatrix} = - \begin{bmatrix} 0 \\ p(X^S) g_\theta(I^S, \theta) \\ 0 \\ \beta^S (p(X^S) g_\theta(I^S,\theta) + (r + \delta - I^S p'(X^S)) g_{I\theta}(I^S, \theta)) \end{bmatrix}$$

Applying Cramer's rule, we have

$$\frac{\partial X^S}{\partial \theta} = \frac{-\delta(r+\delta) p(X^S) \beta^S (p'(X^S) g_\theta(I^S,\theta) + (r+\delta - I^S p'(X^S)) g_{I\theta}(I^S,\theta))}{Det[J_S]} \tag{2.19}$$

$$\frac{\partial I^S}{\partial \theta} = \frac{-\delta(r+\delta)(\delta - I^S p'(X^S)) \beta^S (p'(X^S) g_\theta(I^S,\theta) + (r+\delta - I^S p'(X^S)) g_{I\theta}(I^S,\theta))}{Det[J_S]} \tag{2.20}$$

$$\frac{\partial K^S}{\partial \theta} = \frac{(r+\delta) p(X^S)}{Det[J_S]} \tag{2.21}$$

$$\left(\begin{array}{c} -\beta^S (g(I^S,\theta) p'(X^S) + (\delta - I^S p'(X^S)) g_I(I^S,\theta)) \\ (p'(X^S) g_\theta(I^S,\theta) + (r + \delta - I^S p'(X^S)) g_{I\theta}(I^S,\theta)) \\ -g_\theta(I^S,\theta) \left(\begin{array}{c} -p(X^S) p''(X^S)(u(c^*) - u(m) - e(I^S) + I^S e'(I^S) \\ +\beta^S (g(I^S,\theta) - I^S g_I(I^S,\theta)) + (m - c^*) u'(c^*)) \\ +(\delta - I^S p'(X^S))(r + \delta - I^S p'(X^S))(e''(I^S) - \beta^S g_{II}(I^S,\theta)) \end{array} \right) \end{array} \right)$$

We have previously assumed that autonomy and self-control reinforce each other, so $g_{I\theta}(I^S, \theta) > 0$, and that autonomy on its own has a small positive effect on skills accumulation $g_\theta(I^S, \theta) > 0$. With these assumptions, it is clear that the numerators in each of the equations (2.19)-(2.21) are negative. With the denominator also being negative, $Det[J_S] < 0$, we can see that the effect of a rise in autonomy on the steady state variables X^S, I^S, K^S is positive. ■

Proposition 2.2 states the main result of this version of our model, namely that steady state levels of willpower, self-control and human capital all increase with a rise in autonomy. Recall also that steady state consumption of the rational self is increasing in willpower and human capital, so a rise in autonomy will also increase the steady state consumption. Aside from the upward sloping $\dot{X} = 0$ isocline and the technicalities of the rest of the Assumption 1, the driving force behind this result is the idea that autonomy and self-control reinforce each other in a way that has a beneficial effect on the agent's productivity or human capital. An increase in autonomy then provides an additional incentive for the agent to put in extra self-control effort as this will result in higher human capital, and therefore higher earnings and higher consumption for the rational self. Additionally, the process of putting in extra self-control effort into skill accumulation spills over into increasing the agent's willpower reserves, which in turn reduces the frequency of myopic mistakes in consumption and has an additional positive effect on the consumption of the rational self. Thus, if obesity is inversely related to self-control, a positive relationship between autonomy and self-control, in which a rise in autonomy can increase the marginal product of self-control effort in acquiring skills, can potentially explain the inverse relationship between occupational status and obesity found in the empirical work.

We have so far been assuming that $g_\theta(I^S,\theta) = \varepsilon > 0$, i.e. the effect of autonomy on productivity is small but positive. Technically, this is not necessary for our results, as the assumption of $g_{I\theta}(I^S,\theta) > 0$ is sufficient. Recall that we assumed a positive $g_\theta(I^S,\theta)$ to represent the situation where an increase in flexibility or self-determination has a positive effect on workers' skills. However, it is possible to imagine a situation in which an increase in job autonomy on its own, unaccompanied by an increase in self-control, in fact has a detrimental effect on skills accumulation. For example, if the workers already tend to shirk when unsupervised, an increase in autonomy represented in removal of some supervision could result in workers accumulating fewer skills, and thus becoming less productive. Corollary 2.2 below states that in this case an increase in autonomy can still have a positive effect on steady state levels of willpower, self-control and human capital if the negative first order effect of autonomy is 'not too large'.

Corollary 2.2 *If $g_\theta(I^S,\theta) < 0$, the results of Proposition 2.2 hold if* $-g_\theta(I^S,\theta) < \frac{(r+\delta-I^S p'(X^S))g_{I\theta}(I^S,\theta)}{p'(X^S)}$.

Corollary follows directly from proof of Proposition 2.2.

So far, we have shown how an increase in the degree of job autonomy or self-determination can move the individual to a steady state with higher levels of willpower, self-control and human capital. Part of this effects comes from the positive effect that an increase in autonomy has on earnings, through increase in steady state levels of human capital. Thus, it would not be surprising if an increase in the marginal return to human capital also lead to a higher steady state. We show this below.

Proposition 2.3 *Under Assumptions $1-3$, a rise in the marginal return to human capital, w, increases steady state willpower, X^S, self-control, I^S, and human capital, K^S.*

Proof. Using the same methods as in Proposition 2.2, we arrive at the following results:

$$\frac{\partial X^S}{\partial w} = \frac{-\delta p(X^S)u'(c^*)(p'(X^S)g(I^S,\theta)+(r+\delta-I^S p'(X^S))g_I(I^S,\theta))}{Det[J_S]} \tag{2.22}$$

$$\frac{\partial I^S}{\partial w} = \frac{-p'(X^S)u'(c^*)(p'(X^S)g(I^S,\theta)+(\delta-I^S p'(X^S))g_I(I^S,\theta))(p'(X^S)g(I^S,\theta)+(r+\delta-I^S p'(X^S))g_I(I^S,\theta))}{Det[J_S]} \tag{2.23}$$

$$\frac{\partial K^S}{\partial w} = \frac{-\delta(\delta-I^S p'(X^S))u'(c^*)(p'(X^S)g(I^S,\theta)+(r+\delta-I^S p'(X^S))g_I(I^S,\theta))}{Det[J_S]} \tag{2.24}$$

Since $\delta - I^S p'(X^S) > 0$ by assumption, $Det[J_S] < 0$ as shown in proof of Proposition 2.2, and since all first derivatives in the model are positive, we can see that each of $\frac{\partial X^S}{\partial w}$, $\frac{\partial I^S}{\partial w}$, $\frac{\partial K^S}{\partial w}$ are positive. ■

Proposition 2.3 states that an increase in the marginal return to human capital, increases the steady state levels of willpower, skills and self-control. This result is in parallel to the main result of our first chapter, namely that self-control is increasing with income. However, the mechanism is different here. In our first chapter, higher income led to higher self-control by virtue of reducing the relative payoffs to temptations compared to the underlying wellbeing among the well-off. In this model, an increase in marginal return to human capital incentivises the agent to put in more self-control effort to accumulate skills, as skills now yield a higher return. This translates into higher earnings, both through higher w and K^S, which in turn translate into higher consumption. Through the process of investing self-control effort into skill accumulation, the individual now also accumulates more willpower, which increases the probability of the rational self being in control. Interestingly, the sign of the wage effect is always positive, and does not depend on the cost of self-control. This effect is similar to the efficiency view of wages, in which an increase in the wage rate makes the agent put in more effort, with the addition that in our model this extra effort has beneficial spillover effects on individual's willpower, and further increases consumption.

An alternative view of wages, such as a theory of compensating wage differentials, would suggest that if workers have a preference for autonomy, firms with more routine jobs may have to compensate workers by offering higher wages. In this case we could observe a negative relationship between autonomy and wages in jobs requiring similar levels of human capital. If this is so, does an increase in autonomy still lead to a rise in the steady state? We find that it does as long as the rate of decline of the wage rate with autonomy is smaller than some upper bound, which depends on the total and marginal product of self-control effort and how quickly they rise with autonomy.

Proposition 2.4 *If w is a function of θ, and $w'(\theta) < 0$, an increase in autonomy θ increases the steady state levels of willpower, X, self-control, I, and human capital, K, if*

$$-\frac{w'(\theta)}{w(\theta)} < \frac{p'(X^S)g_\theta(I^S,\theta)+(r+\delta-I^S p'(X^S))g_{I\theta}(I^S,\theta)}{p'(X^S)g(I^S,\theta)+(r+\delta-I^S p'(X^S))g_I(I^S,\theta)} \tag{2.25}$$

Proof. Totally differentiate the steady state equations with respect to θ and solve the resulting system by Cramer's rule to obtain the following results:

$$\frac{\partial X^S}{\partial \theta} = \frac{-\delta p(X^S)}{Det[J_S]}\left(\begin{array}{c} u'(c^*)w'(\theta)(p'(X^S)g(I^S,\theta)+(r+\delta-I^S p'(X^S))g_I(I^S,\theta)) \\ +(r+\delta)\beta^S(p'(X^S)g_\theta(I^S,\theta)+(r+\delta-I^S p'(X^S))g_{I\theta}(I^S,\theta)) \end{array}\right) \tag{2.26}$$

$$\frac{\partial I^S}{\partial \theta} = \frac{-\delta(\delta-I^S p'(X^S))}{Det[J_S]}\left(\begin{array}{c} u'(c^*)w'(\theta)(p'(X^S)g(I^S,\theta)+(r+\delta-I^S p'(X^S))g_I(I^S,\theta)) \\ +(r+\delta)\beta^S(p'(X^S)g_\theta(I^S,\theta)+(r+\delta-I^S p'(X^S))g_{I\theta}(I^S,\theta)) \end{array}\right) \tag{2.27}$$

$$\frac{\partial K^S}{\partial \theta} = \frac{(r+\delta)p(X^S)}{Det[J_S]}\times \tag{2.28}$$

$$\left(\begin{array}{c} u'(c^*)w'(\theta)(p'(X^S)g(I^S,\theta)+(\delta-I^S p'(X^S))g_I(I^S,\theta)) \\ (p'(X^S)g(I^S,\theta)+(r+\delta-I^S p'(X^S))g_I(I^S,\theta)) \\ +(r+\delta) \\ \left(\begin{array}{c} -\beta^S(g(I^S,\theta)p'(X^S)+(\delta-I^S p'(X^S))g_I(I^S,\theta)) \\ (p'(X^S)g_\theta(I^S,\theta)+(r+\delta-I^S p'(X^S))g_{I\theta}(I^S,\theta)) \\ -g_\theta(I^S,\theta)\left(\begin{array}{c} -p(X^S)p''(X^S)(u(c^*)-u(m)-e(I^S)+\beta^S g(I^S,\theta)+(m-c^*)u'(c^*)) \\ +(\delta-I^S p'(X^S))(r+\delta-I^S p'(X^S))(e''(I^S)-\beta^S g_{II}(I^S,\theta)) \end{array}\right) \end{array}\right) \end{array}\right)$$

Consider the expressions for $\frac{\partial X^S}{\partial \theta}$ and $\frac{\partial I^S}{\partial \theta}$ first. With $Det[J_S] < 0$, the term in brackets has to be positive in order for the expressions (2.26) and (2.27) to be positive. This is true when

$$\begin{array}{c} -u'(c^*)w'(\theta)(p'(X^S)g(I^S,\theta)+(r+\delta-I^S p'(X^S))g_I(I^S,\theta)) \\ < (r+\delta)\beta^S(p'(X^S)g_\theta(I^S,\theta)+(r+\delta-I^S p'(X^S))g_{I\theta}(I^S,\theta)) \end{array}$$

Using the fact that $\beta^S = \frac{w(\theta)u'(c^*)}{r+\delta}$ and rearranging, we get:

$$-\frac{w'(\theta)}{w(\theta)} < \frac{p'(X^S)g_\theta(I^S,\theta)+(r+\delta-I^S p'(X^S))g_{I\theta}(I^S,\theta)}{p'(X^S)g(I^S,\theta)+(r+\delta-I^S p'(X^S))g_I(I^S,\theta)} \tag{2.29}$$

Condition (2.29) is also a sufficient condition for $\frac{\partial K^S}{\partial \theta} > 0$, although it is stronger than strictly required. ■

2.2 Welfare

We have now established that a rise in autonomy moves the individual to a higher steady state, with greater levels of willpower, self-control and human capital. Higher human capital translates directly into higher earnings and higher consumption for the rational self, whereas higher willpower allows the rational self to enjoy his consumption more frequently. On the other hand, maintaining this steady state requires higher self-control, and self-control is costly. Hence, the next question is whether the individual is always better off in a higher steady state.

Since in our model the myopic self only exists to represent systematic mistakes in behaviour, we maintain the standard assumption that the individual has a single set of preferences. We measure individual's wellbeing by his experienced utility:

$$V(X, I, K, w, \theta) = \int_0^\infty (p(X)(u(c) - e(I)) + (1 - p(X))u(m))\, e^{-rt} dt \tag{2.30}$$

The steady state welfare is then:

$$V(X^S, I^S, K^S, w, \theta) = p(X^S)(u(c^*) - e(I^S) - u(m)) + u(m) \tag{2.31}$$

Proposition 2.5 *An increase in autonomy, θ, makes the individual better off if:*

$$e'(I^S) < \frac{p'(X^S)(mu'(c^*)+p(X^S)(u(c^*)-e(I^S)-u(m)))}{\delta(p(X^S)-X^S p'(X^S))} + \frac{wg_I(I^S,\theta)u'(c^*)}{\delta} \tag{2.32}$$

Proof. Differentiate the agent's steady state utility with respect to autonomy, θ :

$$\frac{\partial V(X^S,I^S,K^S,w,\theta)}{\partial\theta} = p'(X^S)\frac{\partial X^S}{\partial\theta}(u(c^*) - e(I^S) - u(m)) + p(X^S)(u'(c^*)\frac{\partial c^*}{\partial\theta} - e'(I^S)\frac{\partial I^S}{\partial\theta}) \tag{2.33}$$

Using the fact that $\frac{\partial I^S}{\partial\theta} > 0$ (from Proposition 2.2), re-arrange expression (2.33) to find that $\frac{\partial V(X^S,I^S,K^S,w,\theta)}{\partial\theta} > 0$ if

$$e'(I^S) < \frac{p'(X^S)\frac{\partial X^S}{\partial\theta}(u(c^*)-e(I^S)-u(m))+p(X^S)u'(c^*)\frac{\partial c^*}{\partial\theta}}{p(X^S)\frac{\partial I^S}{\partial\theta}} \tag{2.34}$$

From proof of Proposition 2.2, expressions (2.19)-(2.20), we can see that

$$\frac{\frac{\partial X^S}{\partial\theta}}{\frac{\partial I^S}{\partial\theta}} = \frac{p(X^S)}{(\delta-I^S p'(X^S))} = \frac{p(X^S)^2}{\delta(p(X^S)-X^S p'(X^S))}$$

where the last simplification uses the fact that in steady state, $I^S = \frac{\delta X^S}{p(X^S)}$ (from (2.13)). Notice that since $(\delta - I^S p'(X^S)) > 0$ by assumption, it must be that $(p(X^S) - X^S p'(X^S)) > 0$.

To find $\frac{\partial c^*}{\partial \theta}$, we first use the steady state equations to write c^* as a function of X, and then find $\frac{\partial c^*}{\partial X}$. Thus, using expression (2.14), we can write:

$$
\begin{aligned}
c^* &= m - \frac{m - wK^S}{p(X^S)} \\
&= m + \frac{wg\left(\frac{\delta X^S}{p(X^S)}, \theta\right)}{\delta} - \frac{m}{p(X^S)}
\end{aligned}
\tag{2.35}
$$

and

$$
\frac{\partial c^*}{\partial X} = \frac{mp'(X^S) + w(p(X^S) - X^S p'(X^S)) g_I(I^S, \theta)}{(p(X^S))^2} \tag{2.36}
$$

Then we can write $\frac{\partial c^*}{\partial \theta} = \frac{\partial c^*}{\partial X}\frac{\partial X^S}{\partial \theta}$.

Substituting back into expression 2.34, we have:

$$
e'(I^S) < \frac{p'(X^S)(mu'(c^*) + p(X^S)(u(c^*) - e(I^S) - u(m)))}{\delta(p(X^S) - X^S p'(X^S))} + \frac{wg_I(I^S, \theta)u'(c^*)}{\delta} \tag{2.37}
$$

■

In Proposition 2.5, we show that a small increase in autonomy, and therefore steady state levels of willpower, human capital and self-control, make the individual better off, i.e. increase his experienced utility, provided the marginal cost of self-control at steady state is below an upper bound, defined in (2.37). Increasing the return to human capital, w, or the marginal product of self-control effort, $g_I(\cdot)$, would unambiguously relax this upper bound.

We have established that steady state levels of willpower and self-control rise with income. We have also argued that the ultimate cause of obesity lies with the lack of ability to resist palatable energy-dense foods in the environment of increasing food availability and its falling relative price. Thus, we believe that optimal weight is inversely related to willpower. We can write the weight function for individual i as

$$
W_i = \bar{w}_i + f(X; \omega) \tag{2.38}
$$

where $\bar{w}_i$ is the individual's optimal weight as determined by his physiological characteristics, $f(X;\omega)$ represents the deviation from optimal weight as a function of the amount of willpower, and ω is a vector of exogenous parameters, such as food abundance and price, which can shift the f function. We assume that $f'(X;\omega) < 0$, $f(0;\omega) = w_i^{\max}$ and $\lim_{X\to\infty} f(X;\omega) = w_i^{\min} < 0$, where $w_i^{\max}$ represents the maximum excess weight individual i would achieve in the absence of willpower, and $w_i^{\min}$

is the largest reduction in weight he can survive at if willpower spirals out into the 'control freak' state.

3 Conclusion

In this extension to our baseline income and self-control model, we develop a model in which the degree of job autonomy provides an alternative, and often empirically stronger mechanism through which low socioeconomic status can result in lower self-control and, therefore higher obesity rates. We believe that successful performance at a highly autonomous job requires the kind of human capital, acquiring which demands self-control effort in the first place. Accumulating such human capital would then have spill-over effects on individual's level of willpower. Indeed, we find that an increase in the degree of job autonomy can provide the incentive for an individual to accumulate more human capital, which results not only in higher earnings, but also in greater willpower and self-control.

4 Appendix

Proof of Proposition 2.1

Proof. Step 1: Recall the Hamiltonian and the necessary conditions:

$$\begin{aligned} H(X,K,I,C,\alpha,\beta,\gamma) &= p(X)(u(c)-e(I)-u(m))+u(m)+\alpha(p(X)I-\delta X) \\ &\quad \beta(p(X)g(I)-\delta K)-\gamma(p(X)(c-m)+m-wK) \end{aligned} \tag{4.1}$$

$$\tfrac{\partial H}{\partial c} = p(X)(u'(c)-\gamma)=0 \tag{4.2}$$

$$\tfrac{\partial H}{\partial I} = p(X)(\alpha - e'(I)+\beta g_I(I,\theta))=0 \tag{4.3}$$

$$\dot{\alpha} = (r+\delta-ip'(X))\alpha - p'(X)(u(c)-u(m)-e(i)-\beta g(I,\theta)+(c-m)\gamma) \tag{4.4}$$

$$\dot{\beta} = (r+\delta)\beta - \gamma w \tag{4.5}$$

$$\dot{\gamma} = -\tfrac{\partial H}{\partial Q} = 0 \tag{4.6}$$

$$\lim_{t\to\infty} e^{-rt}H(X,K,I,C,\alpha,\beta,\gamma)=0 \tag{4.7}$$

Define a vector of model parameters and initial conditions $\phi = (\rho, K_0, X_0) = (\delta, r, m, w, K_0, X_0)$. Assume there exists a solution to the necessary conditions $(X^*(t,\phi), K^*(t,\phi), \beta^*$

and the corresponding costate variables $(\alpha(t,\phi), \beta(t,\phi), \gamma(t,\phi)$, with the property that $(X^*(t,\phi), K^*(t,\phi), \beta^*(t,\phi), I^*(t,\phi)) \to (X^S(\rho), K^S(\rho), \beta^S(\rho), I^S(\rho))$ as $t \to \infty$, where $(X^S(\rho), K^S(\rho), \beta^S(\rho), I^S(\rho))$ is the steady state solution of the necessary conditions. Let's show that the necessary transversality condition (4.7) is satisfied. Firstly, γ is a constant so that $\lim_{t\to\infty}\gamma(t) = \gamma$. Second, from condition (4.3), we can see that $\alpha = e'(I) - \beta g_I(I,\theta)$. Because we have assumed that $(X^*(t,\phi), K^*(t,\phi), \beta^*(t,\phi), I^*(t,\phi)) \to (X^S(\rho), K^S(\rho), \beta^S(\rho), I^S(\rho))$ and since $e'(I)$ and $g_I(I,\theta)$ are bounded, we can say that $\alpha(t,\phi) \to \alpha^S(\rho)$ as $t \to \infty$, where $\alpha^S(\rho)$ is the steady state solution for the costate variable. With this in mind, we can confirm that the necessary transversality condition is satisfied by the solution $(X^*(t,\phi), K^*(t,\phi), \beta^*(t,\phi), I^*(t,\phi))$.

$$\lim_{t\to\infty} e^{-rt} H(X^*, K^*, I^*, C^*, \alpha^*, \beta^*, \gamma^*) =$$

$$\lim_{t\to\infty} e^{-rt} \begin{pmatrix} p(X^S)(u(c^S) - e(I^S) - u(m)) + u(m) + \alpha^S(p(X^S)I^S - \delta X^S) \\ +\beta^S(p(X^S)g(I^S) - \delta K^S) - \gamma(p(X^S)(c^S - m) + m - wK^S) \end{pmatrix} =$$

$$\lim_{t\to\infty} e^{-rt} \times const. = 0$$

Next, let's show that the Mangasarain sufficient conditions are satisfied, so that the solution to the necessary conditions is in fact the solution to the rational self's problem. The sufficient conditions are that the Hamiltonian, $H(X,K,I,C,\alpha,\beta,\gamma)$, is concave along the $(X^*(t,\phi), K^*(t,\phi), \beta^*(t,\phi), I^*(t,\phi))$ path, and that $\lim_{t\to\infty} e^{-rt}[\alpha(t,\phi)(X^*(t,\phi) - X(t)) + \beta(t,\phi)(K^*(t,\phi) - K(t)] \leq 0$. Since we have assumed that as $t \to \infty$, $X^*(t,\phi) \to X^S(\rho)$, $K^*(t,\phi) \to K^S(\rho)$, $\alpha(t,\phi) \to \alpha^S(\rho)$, $\beta(t,\phi) \to \beta^S(\rho)$ and because all admissible paths of $X(t)$ and $K(t)$ are bounded, $\lim_{t\to\infty} e^{-rt}[\alpha(t,\phi)(X^*(t,\phi) - X(t)) + \beta(t,\phi)(K^*(t,\phi) - K(t)] = 0$. To check for concavity of the Hamiltonian, we need to show that the Hessian matrix of the Hamiltonian has non-positive eigenvalues. From (4.2) and (4.3), we know that $\gamma(t) = u'(c(t))$ and $\alpha(t) = e'(I(t)) - \beta(t) g_I(I(t),\theta)$ $\forall t \in [0,\infty)$, so the eigenvalues of the Hessian can be written as:

$$\begin{aligned} \nu_1 &= p(X)u''(c^*) \qquad (4.8) \\ \nu_2 &= u'(c^*)w''(K) = 0 \\ \nu_3 &= -p''(X^S)\left[u(m) - u(c^*) + (c^* - m)u'(c^*) + e(I) - Ie'(I) - \beta(g(I,\theta) - Ig_I(I,\theta))\right] \\ \nu_4 &= p(X)\left[\beta g_{II}(I,\theta) - e''(I)\right] \end{aligned}$$

It is clear that $\nu_1 = p(X)u''(c) < 0$ and $\nu_4 = p(X)\left[\beta^S g_{II}(I^S,\theta) - e''(I)\right] < 0$. Concavity of $u(\cdot)$ and $g(I,\theta)$, and convexity of $e(I)$, imply that the second term in ν_3 is negative (we will show this in detail in Step 3 below); therefore, all eigenvalues of the Hessian are non-positive. Since we assumed a linear wage rate for human capital, and so $w''(K) = 0$, one eigen value $\nu_2 = 0$. This implies that the Hamiltonian is not strictly concave and the solution may not be unique.

Step 2. Solve the necessary conditions to find a system of differential equations in (X, K, β, I). Notice that since consumption is constant, we use one of the costate variables to complete the system.

$$\dot{X} = p(X)I - \delta X \tag{4.9}$$

$$\dot{K} = p(X)g(I,\theta) - \delta K \tag{4.10}$$

$$\dot{\beta} = (r+\delta)\beta - wu'(c) \tag{4.11}$$

$$\dot{I} = \frac{1}{-e''(I)+\beta g_{II}(I,\theta)} \left[\begin{array}{c} p'(X)(u(c) - u(m) - e(I) + \beta g(I,\theta) + (m-c)u'(c)) - \\ (r+\delta - Ip'(X))e'(I) + (wu'(c) - \beta I p'(X))g_I(I,\theta) \end{array} \right] \tag{4.12}$$

The steady state occurs when the above system and
$\dot{Q} = -\left[p(X)(c-m) + m - wK\right] e^{-rt}$ are simultaneously equal to zero.

$$p(X^S)I^S - \delta X^S = 0 \tag{4.13}$$

$$p(X^S)g(I^S,\theta) - \delta K^S = 0 \tag{4.14}$$

$$(r+\delta)\beta - wu'(c^*) = 0 \tag{4.15}$$

$$\begin{array}{c} p'(X^S)(u(c^*) - u(m) - e(I^S) + \beta^S g(I^S,\theta) + (m-c^*)u'(c^*)) - \\ (r+\delta - I^S p'(X^S))(e'(I^S) + \beta^S g_I(I^S,\theta)) \end{array} = 0 \tag{4.16}$$

Step 3. Find the Jacobian of the system (4.9)-(4.12) and evaluate at steady state. Then find the eigen values. Due to the size of the Jacobian, it is not practical to present it explicitly in matrix form, but the relevant derivatives are listed below.

$$J = \begin{bmatrix} \frac{\partial \dot{X}}{\partial X} & \frac{\partial \dot{X}}{\partial K} & \frac{\partial \dot{X}}{\partial \beta} & \frac{\partial \dot{X}}{\partial I} \\ \frac{\partial \dot{K}}{\partial X} & \frac{\partial \dot{K}}{\partial K} & \frac{\partial \dot{K}}{\partial \beta} & \frac{\partial \dot{K}}{\partial I} \\ \frac{\partial \dot{\beta}}{\partial X} & \frac{\partial \dot{\beta}}{\partial K} & \frac{\partial \dot{\beta}}{\partial \beta} & \frac{\partial \dot{\beta}}{\partial I} \\ \frac{\partial \dot{I}}{\partial X} & \frac{\partial \dot{I}}{\partial K} & \frac{\partial \dot{I}}{\partial \beta} & \frac{\partial \dot{I}}{\partial I} \end{bmatrix}_{(\dot{X}=0,\ \dot{K}=0, \dot{\beta}=0, \dot{I}=0)} = \tag{4.17}$$

$$\begin{bmatrix} -\delta + I^S p'(X^S) & 0 & 0 & p(X^S) \\ p'(X^S)g(I^S,\theta) & -\delta & 0 & p(X^S)g_I(I^S,\theta) \\ 0 & 0 & r+\delta & 0 \\ \frac{\partial \dot{I}}{\partial X} & 0 & \frac{\partial \dot{I}}{\partial \beta} & \frac{\partial \dot{I}}{\partial I} \end{bmatrix} \tag{4.18}$$

where the last row is

$$\frac{\partial \dot{I}}{\partial X} = \frac{1}{e''(I^S) - \beta^S g_{II}(I^S,\theta)} \times$$

$$\left[p''(X^S) \left[\begin{array}{c} u(m) - u(c^*) + e(I^S) - \beta^S g(I^S,\theta) + \\ (c^* - m)u'(c^*) - I^S e'(I^S) + \beta^S I^S g_I(I^S,\theta) \end{array} \right] \right]$$

$$\frac{\partial \dot{I}}{\partial \beta} = \tfrac{1}{(e''(I^S)-\beta^S g_{II}(I^S,\theta))^2}\times$$

$$\left[\begin{array}{c} (r+\delta-I^S p'(X^S))e'(I^S)g_{II}(I^S,\theta)+ \\ g_I(I^S,\theta)(I^S p'(X^S)e''(I^S)-wu'(c^*)g_{II}(I^S,\theta)) \\ +p'(X^S)\left[(u(m)-u(c^*)+e(I^S)+(c^*-m)u'(c^*))g_{II}(I^S,\theta)-g(I^S,\theta)e''(I^S)\right] \end{array}\right]$$

$$\frac{\partial \dot{I}}{\partial I} = \tfrac{1}{-e''(I^S)+\beta^S g_{II}(I^S,\theta)}\times$$

$$\left[-(r+\delta)e''(I^S)+wu'(c^*)g_{II}(I^S,\theta)+I^S p'(X^S)(e''(I^S)-\beta g_{II}(I^S,\theta))\right]$$

The eigen values, λ, of this jacobian are as follows:

$$\lambda_1 = -\delta \tag{4.19}$$

$$\lambda_2 = r+\delta \tag{4.20}$$

$$\lambda_3 = \tfrac{1}{2(e''(I^S)-\beta^S g_{II}(I^S,\theta))} \tag{4.21}$$

$$\left(\begin{array}{c} re''(I^S)+(\beta\delta-wu'(c^*))g_{II}(I^S,\theta)+ \\ \sqrt{\begin{array}{c}(re''(I^S)+(\beta\delta-wu'(c^*))g_{II}(I^S,\theta))^2+4(e''(I^S)-\beta g_{II}(I^S,\theta)) \\ \left[\begin{array}{c}(\delta-I^S p'(X^S))\left((r+\delta-I^S p'(X^S))e''(I^S)+(\beta^S I^S p'(X^S)-wu'(c^*))g_{II}(I^S,\theta)\right) \\ +p(X^S)p''(X^S)\left(\begin{array}{c}u(m)-u(c^*)+(c^*-m)u'(c^*) \\ +e(I^S)-I^S e'(I^S)-\beta^S(g(I^S,\theta)-I^S g_I(I^S,\theta))\end{array}\right)\end{array}\right]\end{array}} \end{array}\right)$$

$$\lambda_4 = \tfrac{1}{2(e''(I^S)-\beta^S g_{II}(I^S,\theta))} \tag{4.22}$$

$$\left(\begin{array}{c} re''(I^S)+(\beta^S\delta-wu'(c^*))g_{II}(I^S,\theta)- \\ \sqrt{\begin{array}{c}(re''(I^S)+(\beta\delta-wu'(c^*))g_{II}(I^S,\theta))^2+4(e''(I^S)-\beta g_{II}(I^S,\theta)) \\ \left[\begin{array}{c}(\delta-I^S p'(X^S))\left((r+\delta-I^S p'(X^S))e''(I^S)+(\beta^S I^S p'(X^S)-wu'(c^*))g_{II}(I^S,\theta)\right) \\ +p(X^S)p''(X^S)\left(\begin{array}{c}u(m)-u(c^*)+(c^*-m)u'(c^*) \\ +e(I^S)-I^S e'(I^S)-\beta^S(g(I^S,\theta)-I^S g_I(I^S,\theta))\end{array}\right)\end{array}\right]\end{array}} \end{array}\right)$$

To determine the nature of the steady state, we need to ascertain the sign of the eigen values (4.19)-(4.22). Obviously, $\lambda_1 = -\delta < 0$ and $\lambda_2 = r+\delta$. We claim that $\lambda_3 > 0$. Firstly, $e''(I^S) - \beta^S g_{II}(I^S,\theta) > 0$, since $e(I)$, is convex in I, $g(I,\theta)$ is concave in I, and $\beta^S > 0$ if skills are to have value in steady state. Next, $re''(I^S) + (\beta^S\delta - wu'(c^*))g_{II}(I^S,\theta) > 0$ because $(\beta^S\delta - wu'(c^*)) = wu'(c^*)(\frac{\delta}{r+\delta} - 1) < 0$, where we use the fact that $\beta^S = \frac{wu'(c^*)}{r+\delta}$, by equation (4.15). The square root is always positive, making the whole expression for λ_3 positive. Then, λ_4 can only be positive if

$$
\begin{array}{r}
(\delta - I^S p'(X^S))\left[(r+\delta - I^S p'(X^S))e''(I^S) + (\beta^S I^S p'(X^S) - wu'(c^*))g_{II}(I^S,\theta)\right] \\
+p(X^S)p''(X^S)\left[\begin{array}{c} u(m) - u(c^*) + (c^* - m)u'(c^*) + e(I^S) - I^S e'(I^S) \\ -\beta^S(g(I^S,\theta) - I^S g_I(I^S,\theta)) \end{array}\right] \quad < 0
\end{array}
\tag{4.23}
$$

As in the baseline model, we assume that the $\dot{X} = 0$ isocline is upward sloping in the (X, I) plane, which again implies that $\delta - I^S p'(X^S) > 0$. Then it must be that $r + \delta - I^S p'(X^S) > 0$ and $\beta I^S p'(X^S) - wu'(c^*) < 0$. This in turn means that the first term in expression (4.23) is positive. In the second term, $u(m) - u(c^*) + (c^* - m)u'(c^*) < 0$ by concavity of $u(\cdot)$; $e(I^S) - I^S e'(I^S) < 0$ by convexity of $e(I)$; and $-\beta^S(g(I^S,\theta) - I^S g_I(I^S,\theta)) < 0$ by concavity of $g(\cdot)$ in I, and since the costate variable $\beta^S > 0$. Coupled with the fact that $p''(X^S) < 0$ by concavity of $p(X)$, this means that the second term in expression (4.23) is also positive. Therefore, under the assumption of $\delta - I^S p'(X^S) > 0$, condition (4.23) cannot hold. Thus, there are two positive and two negative eigenvalues.

If instead we assume that the $\dot{X} = 0$ isocline is downward sloping in the (X, I) plane, $\delta - I^S p'(X^S) < 0$, it is possible to have a positive λ_4 if

$$
\begin{array}{c}
\left|(\delta - I^S p'(X^S))\left[(r+\delta - I^S p'(X^S))e''(I^S) + (\beta^S I^S p'(X^S) - wu'(c^*))g_{II}(I^S,\theta)\right]\right| > \\
p(X^S)p''(X^S)\left[\begin{array}{c} u(m) - u(c^*) + (c^* - m)u'(c^*) + e(I^S) - I^S e'(I^S) - \\ \beta^S(g(I^S,\theta) - I^S g_I(I^S,\theta)) \end{array}\right]
\end{array}
\tag{4.24}
$$

In this case, there would be only one negative eigen value, $\lambda_1 = -\delta$, and therefore only one path in the (X, K, β, I) space that leads to the steady state. ■

References

[1] Ball, Kylie and David Crawford (2005), "Socioeconomic Status and Weight Change in Adults: A Review", *Social Science and Medicine*, vol 60

[2] Baumeister, R.F., Bratslavsky, E., Muraven, M., & Tice, D.M. (1998), "Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74, 1252-1265.

[3] Baumeister, Roy F. and Kathleen D. Vohs (2007), "Self-Regulation, Ego Depletion and Motivation", *Social and Personality Psychology Compass*, vol (1)

[4] Bernheim, Douglas and Antonio Rangel (2004), "Addiction and Cue Triggered Decision Processes", *American Economic Review*, vol 94 (5), pp. 1558 – 1590

[5] Blanchflower, David G. and Andrew J. Oswald (2004), "Well-Being over Time in Britain and the USA", *Journal of Public Economics*, vol 88 (7-8), pp. 1359 - 1386

[6] Blundell JE & King NA (1996), "Overconsumption as a cause of weight gain: behavioural–physiological interactions in the control of food intake (appetite)," In The Origins and Consequences of Obesity (Ciba Foundation Symposium 201), pp. 138–158. Chichester, UK: Wiley.

[7] Blundell JE, Lawton CL, Cotton JR & Macdiarmid JI (1996), "Control of human appetite: implications for the intake of dietary fat", Annual Review of Nutrition 16, 285–319.

[8] Bray, George A. and Claude Bouchard (2008), *Handbook of Obesity: Clinical Applications*, Taylor and Francis Inc

[9] Caballero B (2007), "The Global Epidemic of Obesity: An Overview," *Epidemiology Review*, 29, pp. 1-5

[10] Department of Health (2006), Health Survey for England

[11] Drewnowski, A. 1997, "Taste preferences and food intake," *Annual Review of Nutrition,* 17: 237-53.

[12] Drewnowski, Adam and SE Spencer (2004), "Poverty and Obesity: The Role of Energy Density and Energy Costs", *The American Journal of Clinical Nutrition*, vol 79, pp. 6-16

[13] Flegal, Katherine M., Carroll, Margaret D., Ogden, Cynthia L., and Johnson, Clifford L., (2002), "Prevalence and Trends in Obesity Among US Adults, 1999-2000," *The Journal of the American Medical Association*, 288 (14), 1728-1732

[14] Friedman M., Reed D. and Mela D, (1992), "Sensory and metabolic influences on fat intake," In *Dietary Fats: Determinants of Preference, Selection and Consumption.* Mela D (Ed.). Elsevier Applied Science, NY, USA, 117–137

[15] Fudenberg, Drew and David K. Levine (2006), "A Dual Self Model of Impulse Control", *American Economic Review*, vol 96 (5)

[16] Fogel, Robert W. (1994) "Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy," *NBER Working Papers 4638*, National Bureau of Economic Research

[17] Graham, Carol and Andrew Felton (2005), "Variance in Obesity Across Cohorts and Countries: A Norms-based Explanation using Happiness Surveys", *CSED Working Paper* No. 42

[18] International Obesity Taskforce (2004)

[19] McLaren, Lindsay (2007), "Socioeconomic Status and Obesity", *Epidemiologic Reviews* vol 29, pp. 29 – 48

[20] Mark Muraven, Dianne M Tice, Roy F Baumeister, (1998), "Self-Control as a limited resource: Regulatory depletion patterns," *Journal of Personality and Social Psychology,* Vol 74 Issue: 3 Pages: 774-789

[21] Muraven, M., Baumeister, R. F., & Tice, D. M. (1999), "Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise," *Journal of Social Psychology,* 139, 446-457

[22] Oaten, M., & Cheng, K. (2006), "Longitudinal gains in self-regulation from regular physical exercise," *British Journal of Health Psychology.*

[23] Oaten, M., & Cheng, K. (2007), "Improvements in self-control from financial monitoring," *Journal of Economic Psychology,* 28, 487-501

[24] Sanchez-Villegas A, Delgado-Rodriguez M, Martinez-Gonzalez MA & De Irala-Estevez J (2003), "Gender, age, socio-demographic and lifestyle factors associated with major dietary patterns in the Spanish Project SUN," *Clinical Nutrition,* 57, 285–292.

[25] Shefrin, H. M. and Thaler R. H. (1981), "An Economic Theory of Self Control," *Journal of Political Economy*

[26] Sobal, J. and Stunkard A. J. (1989), "Socioeconomic Status and Obesity: A Review of the Literature," *Psychological Bulletin,* vol 105 (2), 260-275

[27] Swinburn, B. A., Egger, G. J. and Raza, F. (1999), "Dissecting Obesogenic Environments: The Development and Application of a Framework for Identifying and Prioritising Environmental Interventions for Obesity," *Preventative Medicine,* 29, 563–570.

[28] Ulijaszek, S.J. (2007), "Obesity: A Disorder of Convenience", *Obesity Reviews,* 8 (Suppl.1), 183-187.

[29] Vohs, K.D. & Heatherton, T.F. (2000), "Self-regulatory failure: A resource depletion approach," *Psychological Science,* 11, 249-254.

[30] Wardle, Jane, Jo Waller and Martin J. Darvis (2002), "Sex Differences in the Association of Socioeconomic Status with Obesity", *American Journal of Public Health,* vol 92 (8)

[31] World Health Organization. (2004), "Obesity and overweight," Geneva: WHO

[32] Yanovski, Susan (2003), "Sugar and Fat: Cravings and Aversions," *The Journal of Nutrition,* vol 133, pp. 835-7

Part 3

Not in Front of The Children: A Model of Parental Influence on Self-Control

Abstract

In their "Willpower and Personal Rules", (2004), Benabou and Tirole build a self-signaling model of personal rules based on self-reputation, in which people are uncertain about their underlying willpower type but can infer it from their own past actions. However, in equilibrium of their model, full spectrum of self-control outcomes can be achieved depending on agents' initial beliefs, which remain exogenous. In this chapter, we put their self-signaling model in the dynamic overlapping generations context, which provides a mechanism for the formation of initial beliefs and generates heterogeneous behaviour among agents of the same type driven by different parental choices. We show that, conditional on type, children of parents who exercised more self-control during their lifetime, have higher self-confidence, exercise more self-control themselves and are at least ex ante better off. We find that this heterogeneity persists from two to infinite generations set-up with the long run fraction of population exercising self-control being lower with the influence of parental behaviour than without. Introduction of parental altruism retains the heterogeneity of children's behaviour but also induces parents to exercise more self-control, especially when observed by children in later stages of their life.

1 Introduction

The existence of personal rules, such as diets, dry laws, exercise routines and the daily earning targets of New York City cab drivers[20], and their use as internal methods of self-control pose a curious question for economics – insofar as these rules are purely self-imposed and at least partially successful, how do they work? Consider a dieter who has resolved to never eat dessert at dinner. Imagine that on day one of her diet she is presented with a particularly decadent chocolate fondue. What incentive does she have to stick to her self-imposed diet when this one piece of dessert will make no difference to her weight or waistline? Why not have the dessert "just this once" and resume the diet tomorrow? And if she succumbs to the fondue with this logic today, what is to prevent her from doing the same tomorrow? More generally, in the absence

[20] Anecdotal evidence on personal rules abounds; research evidence includes Camerer (1997) on daily earning rules of New York city cabdriver, and evidence on use of mental accounting rules from Thaler (1980, 1985), Zelizer (1997), Heath and Soll (1996), Americs, Caplin and Leahy (2003)

of external commitment, what provides the incentive for an individual to overcome the pull towards immediate gratification and stick to the ex ante preferred internal rule, when any single deviation would have only a negligible effect on the long term outcome?

In their "Willpower and Personal Rules" (2004), Benabou and Tirole provide an answer to this question with a model of personal rules based on self-reputation. The idea is that, if people are uncertain about their underlying willpower, they may try to infer it from their own past actions, so that future beliefs about the strength of one's will or self-control are directly influenced by current choices, which people may come to regard as "indicative of what kind of person they are". At the same time, belief in own willpower, and therefore ability to resist temptation, is needed in order to attempt self-control in the first place. If the dieter does not believe she would be able to stick to the diet tomorrow, the benefit of sticking to it today is rather reduced. Thus, the motivation to maintain a positive self-image, driven by the preference for self-control in the future, transforms current choices into potential 'precedents' for future behaviour, and it is the fear of creating bad precedents, which diminish one's self-belief, that provides the additional incentive to overcome the bias towards immediate gratification and forego the fondue. Formalized in a self-signaling model of self-control, based on hyperbolic discounting, imperfect knowledge of willpower and imperfect recall, the idea of self-reputation provides a mechanism which can sustain rule-based behaviour. In an important contribution to internal self-control literature, Benabou and Tirole are able to characterise the beneficial "bright line" rules of self-control, as well as the harmful compulsive behaviours, and make such curious predictions as the idea that by inhibiting a build up of self-reputation, an initial period of externally enforced controls can reduce the likelihood of individual attempting self-control in the future. However, the main result of their model, which drives the subsequent predictions, is that self-control is increasing with initial self-confidence, captured by initial beliefs about own willpower. In fact, depending on initial beliefs, a full range of behaviour can be sustained in equilibrium - from no self-control exercised by any type of agent, regardless of underlying willpower type, to at least some self-control attempted by all agents, even the weak-willed ones. Although central to determining the equilibrium outcome, this initial self-confidence remains exogenous throughout the model. Yet, the significance of initial beliefs to determining the outcome raises the question of how people of similar type may come to have different initial beliefs.

In this paper we extend the Benabou and Tirole model by introducing a mechanism which can endogenously generate different initial beliefs among agents of similar type. We put the self-signaling model in the dynamic overlapping generations context, which provides a mechanism for the formation of initial beliefs and generates heterogeneous equilibrium behaviour among agents of the same type, driven by different parental choices. We argue that if individuals are uncertain about their genetic traits, and if such traits can be transmitted from parent to offspring, observing parents' behaviour can reveal information about one's own type. In our set-up, children of those parents who exercised more self-control have higher initial self-confidence than children of less self-controlled parents. Then, conditional on type, this higher self-confidence leads the children of self-controlled parents to exercise more self-control themselves. This heterogeneity in beliefs and behaviour persists from two-generations to infinite-generations set-up. We find that the fraction of children exercising self-control in the long run (infinite generations) equilibrium is lower with observation of parental behaviour than without, due to the fact that failures reveal more information than successes in this self-signaling model. Insofar as more self-control leads to ex ante higher welfare in the Benabou and Tirole model (for all but the weakest types), the children of the more self-controlled parents are on average at least ex ante better off. Introduction of parental altruism retains the heterogeneity of children's behaviour but also induces parents to exercise more self-control when they are being observed by their children. The level of self-control exercised by parents increases further when parents are observed by children in the later stage of their life.

We believe that family background is an important and plausible source of initial self-confidence for two reasons. Firstly, economic status is transmitted from parent to offspring[21]. Bowles and Gintis (2001) show that the extent of intergenerational status transmission is high, with some parent-offspring correlations exceeding Galton's original estimate (two thirds) for height. As well as direct bequests and more relaxed credit constraints, typical mechanisms for the transmission of economic status tend to reflect the combined effects of genetic and cultural transmission of traits which contribute to economic success, such as cognitive functioning, the degree of access to and quality of education, and income-enhancing group memberships. However, bequests affect only a small proportion of the population and the combination of genetic and cultural inheritance, as well as educational attainment, although strongly significant, do not fully explain the degree of wealth, income or social class persistence. Similarly, although intergenerational correlations in IQ tend to be substantial[22], they explain little of the remaining intergenerational status correlations[23]. Mulligan (1997), for instance, shows that a little more than two fifths of the association in status between parents and children remains unexplained after controlling for a number of measures of educational attainment and IQ, as well as standard demographic variables[24].

However, it has been suggested that correlations in personality traits, such as work

[21] This is the case whether status is measured by earnings, income, education, memebership of social class, occupation, or wealth, although the extent of correlation varies somewhat with the type of meaure used. (Bowles and Gintis, 2000)

[22] Black, Deveraux, Salvanes (2008); Bouchard and McGue (1981)

[23] Bowles and Gintis (2000) estimate that one tewentieth or less of the observed intergenerational status transmission is due to genetic inheritance of IQ.

[24] Similarly, Charles and Hurst (2003) find that almost 35% of intergenerational wealth elasticity remains unexplained after controlling for income, asset ownership propensity, education, gifts and expected bequests.

efficacy, planning behaviours and discount rates, could go some way towards addressing the gap in our understanding of status correlations. Knowles and Postelwaite (2005) argue that if differences in savings behaviours among households of similar circumstances can be explained by differences in basic personality traits, such as the discount rates, then one would expect parents and children to share such traits, and so parental savings behaviour should predict children's investment decisions; indeed, they find that parental savings behaviour predicts both education and savings choices of children's households. Insofar as discount rates can be viewed as a measure of self-control ability; intergenerational correlation in discount rates imply correlations in self-control ability itself. This is our second reason for taking an intergenerational approach to self-control. That self-control contributes to economic well-being is uncontroversial: ability to overcome short term impulses in favour of greater orientation towards the future allows the pursuit of longer term goals, improves work ethic and has a positive effect on savings rates. We argue further that correlations in self-control outcomes, and therefore potentially wealth or status, are driven not only by correlations in actual self-control type, but also by children's beliefs about own type as driven by parental choices. Thus, self-control behaviours, which may lead to higher income outcomes among parents, instill better initial self-confidence among children, and in turn lead to better self-control outcomes, and therefore potential status, among the children's generation.

2 The Benabou and Tirole Model

In this section we outline the simplified version of the model as in Benabou and Tirole (2004). We retain only the features necessary to convey the self-signaling intuition, removing the self-serving memory and attribution problems. This framework rests on two fundamental premises: imperfect knowledge of willpower and imperfect recall. In the next section, we shall build on this model by introducing an overlapping generations mechanism which will endogenise initial self-confidence.

An individual lives for two periods, $t = 1, 2$ each of which is divided into two subperiods, $\tau = 1, 2$. At the beginning of each period, i.e. in subperiod 1, the individual is presented with a self-control problem, in which he faces a trade-off between immediate gratification and delayed, willpower-dependent reward. For example, an individual wakes up in the morning and chooses whether to try and stick to a diet, avoid smoking or alcohol for the day, or attempt to complete a piece of work. At this point, the individual must choose between the following options:

1. Indulge in temptation immediately and not even attempt to exercise willpower (NW). This yields a known instantaneous payoff a.

2. Attempt to exercise willpower (W). In this case, at the beginning of subperiod 2 (e.g. the afternoon), the individual begins to experience the 'cravings' cost c of the willpower activity. For instance, he may get hungry, tired or stressed as the day goes by so that avoiding that indulgent snack or drink, or completing the work he started becomes particularly difficult. The individual is now faced with a further self-control problem. He can either persevere with his willpower activity (P) and receive the payoff $B - c$ at the end of period; or he can give up on his attempt at willpower (G), and receive the payoff b, also at the end of the period.

The timing of the individual's choices and payoffs is depicted in Figure 1[25].

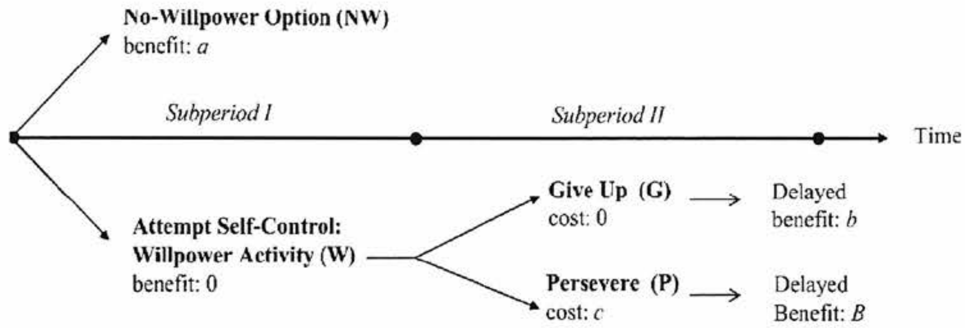


Figure 1. Timing of Payoffs in Benabou and Tirole stage game.

Evaluated ex-ante, it is assumed that willpower and perseverance are the best option, some willpower is the second best option, and no willpower is the worst[26]: $a < b < B - c$. Thus, ex-ante, the optimal choice is clear. What makes this a self-control problem is the agent's discounting of the different payoffs at different points in time: at the time of the decision, the agent tends to overestimate the magnitude of immediate costs and benefits compared to the delayed ones. Specifically, the following quasi-hyperbolic discounting structure is assumed:

- Individual discounts payoffs between periods 1 and 2 at a standard rate δ.
- Faced with immediate temptation payoff a (in every subperiod 1) the individual discounts the future at an additional rate $\gamma < 1$. Faced with an immediate cravings cost c, the individual discounts the future at an additional rate $\beta < 1$, which we refer to as the individual's strength of will throughout this chapter. Further, individual can be of two types: a strong type β_H, or a weak type $\beta_L < \beta_H$.

[25] Diagram taken from Benabou and Tirole (2004) p. 857

[26] Throughout this chapter we focus our attention on the case when even a little self-control is better than none. However, it is possible to conceive of situations where it is best not to start a self-control activity at all, then start and give up (e.g. starting a firm), in which case payoffs to self-control may be non-monotonic: a little self-control is the worst option, none is the second worst and a lot of self-control is still the best.

The above set-up differs from standard models of hyperbolic discounting in two ways. Firstly, β and γ refer to individual's cravings or strength of temptation in different circumstances, and therefore need not be the same. Secondly, Benabou and Tirole assume that whereas γ is known at all times, β is not known ex ante and is "revealed only through the experience of actually putting one's will to test". The justification is that the individual knows his strength of will in resisting temptation in normal times γ, but is unsure about his ability to resist in times of stress, which can be either $\beta_H \leq \gamma$ or $\beta_L < \beta_H$. Indeed, psychological evidence shows that stress, which could be caused by abstinence, proximity of cues which intensify the strength of 'visceral' cravings, or external emotional circumstances can diminish individual's ability to resist temptation. (Loewenstein, 1996, 1999).

The idea that β is not known ex ante - imperfect knowledge of willpower - is the first fundamental premise of this self-signaling model. The second premise is imperfect recall. Not only is β uncertain in advance, it also cannot be recalled through introspection in period 2. It is this uncertainty about one's strength of will, both ex ante and some time after the self-control choice, that allows the individual to infer his type from his actions and provides the incentive for a signaling game between the agent's two temporal selves. Although uncommon in text-book economics, the 'hot-cold empathy gap', whereby the individual struggles to infer from 'cold' introspection the intensity of 'hot' temptation, pain or stress, once again finds support in psychology (Kahneman, Wakker, Sarin, 1997; Loewenstein, Sarin, 1999).

The introduction of the additional discount factors has the following effect on payoffs:

Assumption 1.

(a) *Evaluated ex ante, some self-control is better than none,* $a < b < B-c$ *, but at the time of decision, the individual overestimates the immediate temptation payoff, so that:*

$$b < \frac{a}{\gamma} < B - c \tag{2.1}$$

(b) *If willpower is attempted, then in subperiod* 2*, perseverance* P *is the dominant strategy for the strong types; for the weak types, giving up* G *would be the dominant strategy in the absence of reputational concerns:*

$$\frac{c}{\beta_H} < B - b < \frac{c}{\beta_L} \tag{2.2}$$

(c) *If giving up in period* 1 *meant that at the beginning of period* 2*, no willpower,* NW*, would be chosen with certainty, the weak type would persevere in period* 1*:*

$$\frac{c}{\beta_L} < B - b + \delta(b - a) \tag{2.3}$$

where $\delta(b-a)$ is the maximum reputational stake.

Following Benabou and Tirole, we restrict attention to equilibria which satisfy the assumption of monotonicity in beliefs. That is, recalling perseverance in period one weakly raises the agent's second period beliefs, whereas recalling a lapse of self-control weakly lowers them. In particular, this allows us to rule out the unnatural equilibrium in which perseverance is such bad news that all types choose to give up. Denote by $A^i_{t=1}$ the action taken by agent i in period one. Then, the updated second period beliefs can be written as $\rho_2^P = \Pr(\beta_i = \beta_H \mid A^i_{t=1} = P)$, which is the posterior probability of being a strong type following perseverance, and $\rho_2^G = \Pr(\beta_i = \beta_H \mid A^i_{t=1} = G)$, which is the posterior probability of being a strong type following giving up.

Assumption 2. *Monotonicity in beliefs: $\rho_2^G \leq \rho_1 \leq \rho_2^P$.*

The main result of the Benabou and Tirole model follows below.

Proposition 3.1 *(Benabou and Tirole, Proposition 1, p. 863, without the self-serving memory problem) For $\frac{c}{\beta_L} < B-b+\delta(b-a)$, there is a unique equilibrium. If the agent's initial self-confidence ρ_1 is below a threshold $\rho_1^*(< \rho_2^*)$, he does not put his willpower to test and chooses the NW option. If $\rho_1 > \rho_1^*$, the agent chooses W in the first period, in which case (i) the strong type always perseveres and (ii), the weak type perseveres with probability one for $\rho_1 \geq \rho_2^*$, and with probability $q_1 = \dfrac{\rho_1(1-\rho_2^*)}{\rho_2^*(1-\rho_1)}$ for $\rho_1^* < \rho_1 < \rho_2^*$. In the second period, if G was observed in period 1, NW is chosen with certainty. If P was observed in period 1, W is chosen with probability one for $\rho_1 \geq \rho_2^*$ and with probability $p_2 = \frac{b-B+c/\beta_l}{\delta(b-a)}$ if $\rho_1^* < \rho_1 < \rho_2^*$. In the last subperiod, only the strong type perseveres.*

The intuition of this result is important for our extension so we explain it here. Consider the agent's choice in the last stage of the game, $t = 2$, $\tau = 2$: there are no reputational concerns, so under Assumption 1(b), strong types persevere and weak types give up. With this in mind, consider the agent's problem at the beginning of the second period, $t = 2$, $\tau = 1$: the agent does not recall his type and under Assumption 1(a), only chooses to attempt willpower W if he is sufficiently confident that he would be able to carry it through and persevere at the end of the period. Thus, he attempts willpower if his second period self-confidence, $\rho_2 = \Pr(\beta_i = \beta_H \mid A^i_{t=1})$, is above the threshold ρ_2^*, which solves (2.4) to equate the expected payoffs to W and NW at the time of the decision:

$$\rho_2^*(B-c)+(1-\rho_2^*)b=\frac{a}{\gamma} \tag{2.4}$$
$$\rho_2^*=\tfrac{a/\gamma-b}{B-c-b}$$

Following Benabou and Tirole, we refer to the lower bound restriction on second period beliefs, necessary for the agent to attempt willpower in the second period, as the informativeness constraint:

$$\rho_2=\Pr(\beta_i=\beta_H \mid A^i_{t=1})\geq\rho_2^*=\tfrac{a/\gamma-b}{B-c-b} \tag{2.5}$$

Going back one step further to $t=1$, $\tau=2$, where his type is momentarily revealed to the agent: given Assumption 1(b) and monotonicity of beliefs, it is the dominant strategy for the strong type to persevere. For the weak type, the problem is more complex. On the one hand, he is tempted to give up when experiencing the discounted cravings cost, $\frac{c}{\beta_L}$, but on the other, period two self-confidence provides the motivation for attempting willpower at the beginning of the second period, which is ex ante desirable. As such, the period 2 self-belief is an asset worthy of protection in period 1. Suppose that in equilibrium, W is chosen at $t=2$, $\tau=1$ with probability $0<p_2\leq 1$ if P was played in the first period, and with zero probability otherwise. Then the weak type would choose to persevere if

$$b-(B-\tfrac{c}{\beta_L})\leq\delta p_2(b-a) \tag{2.6}$$

The left hand side of (2.6) is the disutility of resisting temptation at $t=1$, $\tau=2$, and the right hand side is the gain in expected utility in period two from having higher self-confidence. Conditional on initial beliefs, three types of equilibrium outcomes could now arise.

Pooling on willpower and perseverance. Suppose the weak type were to persevere at $t=1$, $\tau=2$ with probability one. Given that the strong type always perseveres, this leaves the period two posterior beliefs unchanged, $\rho_2=\rho_1$. This can only be an equilibrium if the initial self-confidence is already sufficiently high to induce the agent to choose willpower in period two, i.e. the initial beliefs already satisfy the informativeness constraint, $\rho_1\geq\rho_2^*$. Otherwise, if all types were to persevere in period one but initial beliefs were low, $\rho_1<\rho_2^*$, no-one would choose W in period two, which cannot be optimal for the weak type, who only perseveres in the first period to gain confidence to attempt self-control in the second period. Since $\rho_1^*<\rho_2^*$, W is also the optimal choice at the beginning of the first period when $\rho_1\geq\rho_2^*$.

Semi-Separation. With $\rho_1^*<\rho_1<\rho_2^*$, the weak type has the incentive to partially pool with the strong type at $t=1$, $\tau=2$, but with low enough probability to keep his period two self-confidence high. In particular, the weak type will persevere with probability q_1 which makes observing P sufficiently good news to raise the updated period two beliefs to ρ_2^*. By Bayesian updating, his second period beliefs will be:

$$\Pr(\beta_i = \beta_H \mid A_{t=1} = P) = \frac{\rho_1}{\rho_1 + q_1(1-\rho_1)} = \rho_2^* \tag{2.7}$$

Solving (2.7) for q_1, we have $q_1 = \dfrac{\rho_1(1-\rho_2^*)}{\rho_2^*(1-\rho_1)}$. By randomizing with probability q_1 between P and G in the first period, the agent makes his second period self indifferent between W and NW, but for the agent to randomise in the first period, he must be indifferent between P and G in the first place. Thus, the second period self randomises with probability p_2 which solves (2.6) with equality to achieve this indifference:

$$p_2 = \frac{b - B + \frac{c}{\beta_L}}{\delta(b-a)} \tag{2.8}$$

In this case, at the beginning of the game, $t = 1$, $\tau = 1$, the agent would choose W over NW if $\rho_1 > \rho_1^*$ which satisfies:

$$\begin{aligned} &\rho_1^*(B - c + \delta(p_2(B-c) + (1-p_2)a))+ \\ &(1-\rho_1^*)(b + \delta(p_2 b + (1-p_2)a) = \tfrac{a}{\gamma} + \delta a \end{aligned} \tag{2.9}$$

The left hand side of (2.9) is the ex ante expected payoff to choosing W initially and playing the equilibrium strategy described above thereafter, given that the agent believes to be a strong type with probability $\rho_1^*(< \rho_2^*)$; and the right hand side gives the payoff to playing NW in both periods. The latter option does not put the agent's willpower to test and hence leaves the prior beliefs unchanged in period 2. Condition (2.9) simplifies to:

$$\rho_1^* = \frac{a/\gamma - b - \delta p_2(b-a)}{(B-c-b)(1+\delta p_2)} < \rho_2^* \tag{2.10}$$

Pooling on NW. This occurs if $\rho_1 < \rho_1^*$, in which case initial self-confidence is so low that no agent tries to put his willpower to test and $\{NW; NW\}$ is played.

Although the above equilibrium is unique in terms of strategy given beliefs, depending on the initial beliefs, three outcomes can arise. With high initial self-confidence, all types persevere in the first period and choose willpower in the second period; for medium range of initial self-confidence, the strong types always persevere, the weak types sometimes persevere, and every type randomises on willpower in period two; for low values of initial self-confidence, no-one even attempts willpower. This dependence of equilibrium outcomes on initial beliefs raises the question of how these beliefs may come about. We put forward an explanation based on building an overlapping generations version of this model.

3 Continuum of Types amendment of the Benabou and Tirole Model

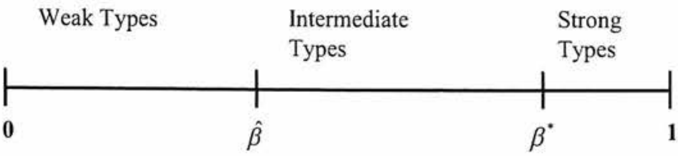
Before proceeding to overlapping generations, we introduce an important amendment to the Benabou and Tirole model in assuming a continuum of agents on the interval $(0,1]$, who differ by the magnitude of their additional (hyperbolic) discount factor β. Specifically, rather than just having two types, β_H and β_L, we assume that an agent's type β is drawn from a continuous distribution $F(\beta)$, which has full support in the interval $(0,1]$[27]. Each agent knows the distribution of types in the population but not his own type. Agents play the stage game as in Benabou and Tirole (2004), described in Section 2. We will show that this amendment (partially) purifies the self-signaling equilibrium, simplifies it and reduces the counter-intuitive mixed-strategy specific effects on the predictions of the overlapping generations model when parental altruism is introduced.

When introducing a continuum of types into a model where the self-control choice is essentially discrete, a new approach to classifying agents into groups which perform similar actions is needed[28]. We call agent i a strong type if he perseveres in the last subperiod, i.e. if β_i is such that $B - \frac{c}{\beta_i} > b$. Then, all agents with $\beta_i \geq \beta^*$, where $\beta^* = \frac{c}{B-b}$, are of strong type. With monotonic beliefs, it remains the dominant strategy for the strong types to persevere in the first period, too. We call agent i an intermediate type if $\beta_i < \beta^*$ *and* he perseveres in the first period when this leads to willpower being chosen with probability one in the second period (and when giving up leads to no willpower with certainty). For such an agent, choosing G at $t = 1, \tau = 2$ yields $b + \delta a$, and choosing P yields $B - \frac{c}{\beta_i} + \delta b$; thus all agents with $\beta_i \in [\hat{\beta}, \beta^*)$ are of intermediate type, where $\hat{\beta} = \frac{c}{B-b+\delta(b-a)} < \beta^*$ is the last intermediate agent who perseveres at $t = 1, \tau = 2$ (agent $\hat{\beta}$ is indifferent between P and G at $t = 1$, $\tau = 2$). All agents with $\beta_i < \hat{\beta}$ cannot be induced to persevere even by the prospect of willpower with certainty in the next period (the maximum reputational stake). We call such agents the weak types. We shall refer to an equilibrium in strategies, which assign different actions to the intermediate types and the weak types when their type is revealed to them in the first period, as a semi-separating equilibrium[29].

[27] To avoid division by zero, we simply assume that the discount factor zero is not included in the support. This also rules out the extreme type of agent who, faced with a cravings cost, attaches no weight at all to the future.

[28] In the Benabou and Tirole (2004) model, action can only be a step-function of type, so the classification into groups adds convenience.

[29] To the extent that we have three types of agents but only two possible actions at any information set, an equilibrium of this model could never be completely separating. Since we are interested in inducing agents other than the strong types to exercise self-control, we focus on the semi-separating equilibriun in which the intermediate types pool with the strong types by persevering.



Denote by $A^i_{t=j,\tau=k}$ the action chosen by agent $i \in (0,1]$ in period $j \in \{1,2\}$, subperiod $k \in \{1,2\}$. Then, for any agent i, a strategy $S_i = \{A_{t=1,\tau=1}(\rho^i_1), A_{t=1,\tau=2}(\rho^i_1, \beta_i), A_{t=2,\tau=1}(\rho^i_2), A_{t=2,\tau=2}(\beta_i)\}$ assigns an action for every information set in the game, conditional on the agent's beliefs and (sometimes) type. At every subperiod $\tau = 1$, the agent does not know his type and, for a given set of model parameters, his action is a function of his beliefs alone; at $t = 1$, $\tau = 2$, his type is momentarily revealed to the agent and his action is a function both of his beliefs and his revealed type; at $t = 2$, $\tau = 2$, action is a function of type only. Although the game is not between different individuals but rather between one agent's two temporal selves, the actions of other types of agents have an effect on agent $i's$ choices insofar as they affect his beliefs regarding the distribution of types in the population.

Proposition 3.2 *Consider the stage game of Benabou and Tirole (2004) with a continuum of types. Under Assumption* 1(*a*) *and monotonicity of beliefs, a semi-separating equilibrium in pure strategies exists if and only if the distribution of types in the population,* $F(\beta)$, *satisfies*

$$(1-F(\beta^*))(B-c+\delta(B-c))+(F(\beta^*)-F(\hat{\beta}))(B-c+\delta b)+F(\hat{\beta})(b+\delta a) \geq \tfrac{a}{\gamma}+\delta a \quad (3.1)$$

and

$$\tfrac{1-F(\beta^*)}{1-F(\hat{\beta})} \geq \rho^*_2 \quad (3.2)$$

In this equilibrium, the following strategy profile S^* *is played:*

- *At* $t = 1$, $\tau = 1$: *Attempt willpower.*
- *At* $t = 1$, $\tau = 2$: *Persevere if* $\beta_i \geq \hat{\beta} = \frac{c}{B-b+\delta(b-a)}$. *Give up otherwise.*
- *At* $t = 2$, $\tau = 1$: *If perseverance was observed at* $t = 1$, $\tau = 2$, *choose willpower with probability* 1. *If giving up was observed, choose no willpower.*
- *At* $t = 2$, $\tau = 2$: *Persevere if* $\beta_i \geq \beta^*$. *Give up otherwise.*

If condition (3.1) does not hold, no agent attempts willpower in the first period and there is a pooling equilibrium on NW.

Proof. The proof is analogous to that of Proposition 1 with the amendments for continuity of types, and can be found in Appendix A. ■

In Proposition 3.2 we construct a semi-separating equilibrium in pure strategies, in which the strong types always persevere, the intermediate types persevere in the first period but not the second and the weak types never persevere. In the second period, the strong types and the intermediate types attempt willpower with probability one, and the weak types with probability zero. Thus, in Proposition 3.2, we have purified the original Benabou and Tirole result by introducing a continuum of agents on the interval $(0, 1]$. To ensure existence of this equilibrium, conditions (3.1) and (3.2) must hold. Condition (3.1) states that the expected payoff to attempting willpower in the first period (LHS), given the agents continue to play according to the equilibrium strategy, must be at least as large as the expected payoff to no willpower (RHS). This is necessary for a separating equilibrium; if initial beliefs were too low to attempt willpower, all agents would pool on NW. Condition (3.1) is therefore a restriction on the distribution of types, which ensures that the initial beliefs are sufficiently favourable for all agents to attempt willpower in the first period. However, it is easy to see that condition (3.1) can be satisfied by any distribution which does not attach disproportionate amount of weight to the bottom end of the support[30]. Condition (3.2) is the informativeness constraint, which ensures that the updated second period beliefs following perseverance are sufficient to induce agents to attempt willpower. Otherwise, the intermediate types would not have the sufficient incentive to pool with the strong types by persevering, so that S^* cannot be an equilibrium.

In fact, Proposition 3.2 characterises the only possible pure semi-separating equilibrium in which the intermediate types pool with the strong types. To see this, suppose condition (3.2) does not hold but there exists some strategy profile S', according to which all agents with $\beta_i \geq \beta'(> \hat{\beta})$ persevere at $t = 1$, $\tau = 2$, where β' solves $\frac{1-F(\beta^*)}{1-F(\beta')} = \rho_2^*$. Then in period 2, upon observing perseverance, agents would be sufficiently confident to choose willpower. However, consider some agent j with $\hat{\beta} \leq \beta_j < \beta'$, who according to S', does not persevere at $t = 1$, $\tau = 2$. This agent would have an incentive to deviate: if he chose P at $t = 1$, $\tau = 2$, he would gain sufficient confidence to choose willpower at $t = 2$, $\tau = 1$ given that beliefs are formed according to S' (even though these beliefs are now incorrect), which is a profitable deviation as long as $B - \frac{c}{\beta_j} + \delta b \geq b + \delta a$, which is true for all $\beta_j \geq \hat{\beta}$. Thus, all agents with $\hat{\beta} \leq \beta_j < \beta'$ would have an incentive to deviate and so S' cannot be an equilibrium.

[30] By Assumption 1(a), $B-c+\delta(B-c) > a/\gamma+\delta a$ and $B-c+\delta b > a/\gamma+\delta a$, and only $b+\delta a < a/\gamma+\delta a$. Threfore, although specific examples would depend on model parameters, in principle any distribution which does not attach "too much" weight to the interval $(0, \dfrac{c}{B-b+\delta(b-a)})$ would satisfy condition 3.1.

Although there cannot be another pure semi-separating equilibrium when the informativeness constraint in condition (3.2) does not hold, there can nevertheless be a mixed one.

Proposition 3.3 *Consider the stage game of Benabou and Tirole (2004) with a continuum of types. Under Assumption* 1(*a*) *and with monotonic beliefs, a mixed strategy semi-separating equilibrium exists if:*

$$(1-F(\beta^*))(B-c+\delta(p_2(B-c)+(1-p_2)a))+(F(\beta^*)-F(\beta'))(B-c+\delta(p_2b+(1-p_2)a)) \tag{3.3}$$

$$+F(\beta')(b+\delta a)\geq \tfrac{a}{\gamma}+\delta a$$

and

$$\tfrac{1-F(\beta^*)}{1-F(\beta')}=\rho_2^* \tag{3.4}$$

for some β'*, where* $\hat{\beta}<\beta'<\beta^*$.

In this equilibrium, the following strategy profile S^{m*} *is played:*

- *At* $t=1$, $\tau=1$ *: Attempt willpower.*
- *At* $t=1$, $\tau=2$ *: Persevere if* $\beta_i\geq\beta'$*, where* β' *solves* $\frac{1-F(\beta^*)}{1-F(\beta')}=\rho_2^*$*. Give up otherwise.*
- *At* $t=2$, $\tau=1$ *: If perseverance was observed at* $t=1$, $\tau=2$*, choose willpower with probability* $p_2=\frac{b-B+c/\beta'}{\delta(b-a)}$*. If giving up was observed, choose no willpower.*
- *At* $t=2$, $\tau=2$ *: Persevere if* $\beta_i\geq\beta^*$*. Give up otherwise.*

Moreover, when condition (3.3) holds with strict inequality , this mixed strategy equilibrium is unique.

Proof. The proof is analogous to that of Proposition 3.2 and can also be found in Appendix A. ■

Proposition 3.3 describes the mixed equilibrium, which is the closest equivalent of the original Benabou and Tirole result in the continuum of agents setting. In this mixed equilibrium, agents randomise only in period 2: upon observing perseverance, willpower is chosen with probability $p_2<1$. Compared to the pure equilibrium, this reduces the reputational stake in period 1, and therefore reduces the incentive for the intermediate types to persevere. Therefore, the last agent to persevere at $t=1$, $\tau=2$ in the mixed equilibrium is of type $\beta'>\hat{\beta}$, i.e. fewer intermediate type agents persevere in the mixed equilibrium than in the pure. In fact, for agents to randomise in period 2, the last agent to persevere in period 1 must make all second period selves indifferent between willpower and no willpower in period 2, so that the informativeness constraint holds with equality (hence condition (3.4)). Similarly, in period 2, agents must randomise between W and NW with exactly the probability which makes agent β' indifferent between P and G , i.e. $p_2 = \frac{b-B+c/\beta'}{\delta(b-a)}$.

Similar to condition (3.1), condition (3.4) ensures that initial self-confidence is sufficiently high for agents to attempt willpower at the beginning of the game, given S^{m*} is played. When (3.4) holds with strict inequality, this is the only mixed equilibrium. To see this, consider what happens if agents randomise between persevering and giving up in the first period. The strong types and the weak types continue to play their dominant strategies and will not randomise, but suppose there exists an equilibrium in which the intermediate types randomise between P and G at $t = 1$, $\tau = 2$. Since for all $\beta_i \geq \hat{\beta}$, P yields a higher payoff than G if it's followed by W with certainty, the middle types would also have to randomise at $t = 2$, $\tau = 1$ to make themselves indifferent between P and G in the first period. Thus, we could imagine that the middle types randomise with such a probability $q_1(\beta_i)$ at $t = 1$, $\tau = 2$ which makes the second period self is indifferent between W and NW[31], and with some probability $p_{2i} = \dfrac{b - B + \frac{c}{\beta_i}}{\delta(b-a)}$ at $t = 2$, $\tau = 1$, which makes the agent indifferent between P and G. The second period randomisation probability p_{2i} must depend on β_i, since for every type, there would be a different probability which makes him indifferent between P and G at $t = 1$, $\tau = 2$. However, at $t = 2$, $\tau = 1$, the agent cannot recall his type, and therefore cannot condition his second period actions on his type β_i. Thus, any strategy which involves the middle types randomising at $t = 1$, $\tau = 2$, cannot be an equilibrium.

So far, we have established that two different types of separating equilibria can occur depending on the distribution of types. When there is a relatively large mass at the top of the types distribution and model parameters are such that it is possible to have $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} > \rho_2^*$, the pure equilibrium of Proposition 3.2 is played. When the distribution of types given the model parameters is not sufficiently favourable for the existence of pure semi-separating equilibrium, $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} < \rho_2^*$, a mixed equilibrium could still be played as long as there exists some β' such that $\frac{1-F(\beta^*)}{1-F(\beta')} = \rho_2^*$. Due to the reduced reputational stake, the last agent to persevere in the mixed equilibrium of Proposition 3.3 is of stronger type than the last agent to persevere in the pure equilibrium of Proposition 3.2, $\beta' > \hat{\beta}$. It follows that the pure and the mixed equilibria do not overlap: if $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} > \rho_2^*$, there can only be a pure equilibrium (since in this case it is not possible to have $\frac{1-F(\beta^*)}{1-F(\beta')} = \rho_2^*$ with $\hat{\beta} < \beta'$); conversely, we have already shown that there can be no pure equilibrium when $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} < \rho_2^*$. Finally, when the distribution of types is so unfavourable that even condition (3.4) does not hold, there can only be a pooling equilibrium on NW.

[31] For instance, the intermediate types could randomise with probability $q_1(\beta_i)$ at $t = 1$, $\tau = 2$ which solves $\frac{1-F(\beta^*)}{1-F(\beta^*)+(F(\beta^*)-F(\hat{\beta}))\int_{\hat{\beta}}^{\beta^*} q_1(\beta_i)f(\beta_i)d\beta / \int_{\hat{\beta}}^{\beta^*} f(\beta_i)d\beta} = \rho_2^*$.

To the extent that the initial distribution of types corresponds to the initial beliefs in the one generation case, our findings with a continuum of types so far mirror the original results of Benabou and Tirole: the degree of self-control exercised in equilibrium is increasing with initial beliefs, which are determined by the initial distribution of types. However, we have partially 'purified' the original Benabou and Tirole model, having established the existence of a pure strategy semi-separating equilibrium. In the remainder of this paper, we shall focus on the pure equilibrium wherever possible, however, the existence of two different equilibria depending on initial beliefs, will become useful when we come to describe potential differences in the behaviour of children, depending on the amount of self-control exercised by their parents.

4 Overlapping Generations Model with a Continuum of Types

In the Benabou and Tirole (2004) model, described above, initial beliefs are critical in determining the self-control outcome. We propose to endogenise these beliefs by putting the basic model into an intergenerational set-up. Our extension is based on the same premises as the original Benabou and Tirole, namely imperfect knowledge of own willpower and imperfect recall, plus one more - a degree of genetic heritability of traits. In an environment where underlying traits are uncertain but actions are observable, genetic inheritance implies that children's expectations about their own type can be partially based on observing their parents' behaviour.

Suppose now that populations can reproduce. Reproduction is asexual so that each family consists of one parent and one child, and each generation is a continuum of agents on $(0, 1]$. Borrowing from the learning literature, we assume the following inheritance (mutation) mechanism: with probability θ, the child inherits exactly his parent's type; with probability $1-\theta$ his type is a random draw from the distribution of types of the parent population. This structure implies that there is a probability mass on the child being very similar to the parent, but the child could also mutate into any other type in the population. Note that under this assumption, the distribution of the parents' types and the children's types will be the same[32].

We also assume that agents now live for three periods: in periods 1 and 2 they play the stage game described above, but there is an additional period 0 in which children observe the actions of their parents. Children's period 0 coincides with parents' period

[32] Given the inheritance mechanism, the distribution of children's types $G(\beta)$ will be : $G(\beta) = \Pr(\beta_{child} \leq \beta) = \theta F(\beta) + (1-\theta)F(\beta) = F(\beta)$. Intuitively, what is the probability that a child's type is below some β'? Either the child's parent was himself a type below β' and the child inherited his type, or the parent was a type above β' and the child did not inherit his type but rather was drawn on the part of the parental distribution below β'.

1. Children do not know their own nor their parents' type but infer it from behaviour. Children form initial beliefs about their own type through Bayesian updating, by combining the information they have on their parents' choices and the distribution of types (and strategies) in the population as a whole. To start with, we assume that children observe only the first period actions of their parents, and not the second. The intuition is that children are exposed to parents' choices while they live at home, but once they move out, parents' actions become more difficult to observe. We shall consider the effects of relaxing this condition in the next section.

Assumption 3. *(a) Any generation $k \in \{2, 3, ...\}$ directly observe $t = 1$, $\tau = 2$ actions of their parents only. (b) The equilibrium strategies of all generations are common knowledge*[33].

4.1 Two Generations

We start by considering the case of just two generations: the first generation of parents and the second generation of children. For the moment, we assume that parents derive no additional utility from the welfare of the child - there is no parental altruism. We investigate how the information revealed through parents' choices can lead to different initial self-confidence, and therefore different self-control choices, among their children.

With no parental altruism, the behaviour of the parents' generation is identical to that of the single agent, described in Propositions 3.2 and 3.3 of the previous section. To see how differences in parents' behaviour can affect children's choices, we need heterogeneity of behaviour on the parents' side in the first place, and therefore assume that parents play a semi-separating equilibrium, i.e. the distribution of types is sufficiently high for parents to attempt willpower in the first period. Further, for the purpose of finding the purest equilibrium possible, we assume that the semi-separating equilibrium of the parents' generation is pure. As will become clear shortly, this assumption does not fundamentally affect children's behaviour, but it will become important once parental altruism is introduced.

Assumption 4. *Parents' generation play the pure semi-separating equilibrium S^* of Proposition 2, and therefore:*

(a) $(1-F(\beta^*))(B-c+\delta(B-c))+(F(\beta^*)-F(\hat{\beta}))(B-c+\delta b)+F(\hat{\beta})(b+\delta a) \geq \frac{a}{\gamma}+\delta a$;

(b) $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} \geq \rho_2^*$.

[33] Since in this model the game is not between different agents but rather between agents' temporal selves, the extra assumption on common knowledge of other agents' strategies is needed to enable agents to make inference about their parents', and hence their own, type.

With this in mind, let's consider how the second generation's initial beliefs are formed. Since children know their parent's first period choices but not their type, they must make the same inference about their parent's type as the parent's second period self. Denote by $\rho_1^{2|P}$ and $\rho_1^{2|G}$ initial beliefs of the second generation children whose parents persevered and gave up, respectively.

Lemma 3.1 *Children of persevering parents have higher initial self-confidence than children of parents who had given up:* $\rho_1^{2|P} \geq \rho_1^{2|G}$.

The lemma follows immediately from the derivation of children's' beliefs. Denote by β_{ki} the type of the individual from dynasty i in generation k, where $k = 1, 2$. Further, denote by $A^{1i}_{t=1,\tau=2}$ the action selected by agent $i's$ parent at $t = 1$, $\tau = 2$. Under Assumption 4, if agent i observed his parent persevere, he knows his parent is a strong type with probability $\frac{1-F(\beta^*)}{1-F(\hat{\beta})}$. Then, agent $i's$ probability of being a strong type is the sum of the probability of inheriting his parent's type in the case that the parent was in fact a strong type, plus the probability that he is drawn on the high part of the distribution. Agent $i's$ initial beliefs, conditional on his parents persevering, can be written as:

$$\begin{aligned}\rho_1^{2|P} &= \Pr(\beta_{2i} \geq \beta^* \mid A^{1i}_{t=1,\tau=2} = P) = \theta\frac{1-F(\beta^*)}{1-F(\hat{\beta})} + (1-\theta)(1-F(\beta^*)) \\ &= \frac{1-F(\beta^*)}{1-F(\hat{\beta})}(1-(1-\theta)F(\hat{\beta}))\end{aligned} \tag{4.1}$$

If, on the other hand, agent i observed his parent give up, he knows with certainty his parent is not a strong type, and so the only hope for agent i himself to be a strong type is if he is randomly drawn on the high part of the distribution.

$$\rho_1^{2|G} = \Pr(\beta_{2i} \geq \beta^* \mid A^{1i}_{t=1,\tau=2} = G) = (1-\theta)(1-F(\beta^*)) \tag{4.2}$$

It is clear from (4.1) and (4.2) that $\rho_1^{2|G} < 1 - F(\beta^*) < \rho_1^{2|P}$ for all $\theta \neq 0$ and $\hat{\beta} < \beta^*$.

We are interested in how this difference in initial beliefs due to observing parents' choices and some degree of heritability, can affect children's behaviour. In the first period, it remains the dominant strategy for the strong types to persevere, so their choices are not directly affected by parents' behaviour, provided their initial self-confidence is not lowered to the extent that they do not even put their willpower to test. It also remains the case that the very weak types with $\beta_i < \hat{\beta}$ cannot be induced to persevere no matter how high their initial beliefs are. However, the first period behaviour of the intermediate types, and in particular their choice between persevering and giving up, strongly depends on their initial beliefs and can therefore be affected by parents' behaviour. In addition, parental choices have the potential to affect second period behaviour of both the strong and the intermediate types by influencing whether these agents have sufficient self-confidence to attempt willpower with any probability.

Proposition 3.4 *Under Assumptions* 3 *and* 4, *and with monotonicity of beliefs, there exists a semi-separating equilibrium in pure strategies for generation* 2. *In this equilibrium, children of persevering parents play the following strategy* S_2^P*:*

- *At* $t = 1$, $\tau = 1$: *Attempt willpower.*
- *At* $t = 1$, $\tau = 2$: *Persevere if* $\beta_{2i} \geq \hat{\beta}$. *Give up otherwise.*
- *At* $t = 2$, $\tau = 1$: *Choose willpower with probability* 1 *if perseverance was observed in the previous period; otherwise, choose no willpower.*
- *At* $t = 2$, $\tau = 2$: *Persevere if* $\beta_{2i} \geq \beta^*$. *Give up otherwise.*

Children of giving-up parents play the following strategy S_2^G :

- *At* $t = 1$, $\tau = 1$: *Attempt willpower if*

$$\begin{aligned}(1-\theta)(1-F(\beta^*))(B-c+\delta p_2(B-c) \\ +(1-p_2)a)+(1-\theta)(F(\beta^*)-F(\beta''))(B-c+\delta(p_2 b+(1-p_2)a)) \\ +[\theta+(1-\theta)F(\beta'')](b+\delta a) \geq \tfrac{a}{\gamma}+\delta a\end{aligned} \tag{4.3}$$

- *At* $t = 1$, $\tau = 2$: *Persevere if* $\beta_{2i} \geq \beta^*$. *Persevere if* $\beta_{2i} \geq \hat{\beta}$ *and* $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \geq \rho_2^*$, *or if* $\beta_{2i} \geq \beta''$ *and if* $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} < \rho_2^*$ *but* $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\beta'')} = \rho_2^*$ *for some* $\beta'' \in (\hat{\beta}, \beta^*)$. *Give up otherwise.*
- *At* $t = 2$, $\tau = 1$: *Choose willpower with probability* 1 *if perseverance was observed in the previous period and* $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \geq \rho_2^*$*; Choose willpower with probability* p_2 *if willpower was observed in previous period and* $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\beta'')} = \rho_2^*$. *Otherwise, choose no willpower.*
- *At* $t = 2$, $\tau = 2$: *Persevere if* $\beta_{2i} \geq \beta^*$. *Give up otherwise.*

Proof of Proposition 3.4 can be found in Appendix A.

In Proposition 3.4, we show that heterogeneous equilibrium behaviour, driven by different parental choices, is possible among children of the same willpower type. In equilibrium of Proposition 3.4, children of persevering parents play the same pure strategy as the first generation - although these children have higher initial beliefs than their parents, they cannot do better than their parents as the self-control choice is discrete and, by assumption, parents already play the pure strategy that yields the highest self-control outcome. Children of giving up parents can also play this pure strategy but only as long as their updated second period beliefs satisfy the informativeness constraint, $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \geq \rho_2^*$, which would be more binding for the children of giving up parents as their initial self-confidence is lowered by observation of low self-control behaviour among their parents. Then, if the distribution of types in the population is not sufficiently favourable, and/or the degree of type heritability is large, the pure strategy informativeness constraint may not be met for the children of giving up parents, inducing them to play the mixed strategy instead.

In the latter case, when $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} < \rho_2^*$, conditional on type, children of giving up parents exercise less self-control than children of persevering parents in equilibrium of .Proposition 3.4. This effect is two-fold. First, fewer children of giving up parents persevere in the first period: the weakest child of persevering parent to persevere is $\hat{\beta}$, whereas the weakest child of giving up parent to persevere is $\beta'' > \hat{\beta}$. Second, conditional on persevering in the first period, children of persevering parents choose willpower with probability one in the second period, whereas children of giving up parents only choose willpower with probability $p_2 < 1$. Thus, a fraction of children of giving up parents with $\beta_{2i} \in [\hat{\beta}, \beta'')$ do not persevere in equilibrium even though they would have, had they not observed their parents' low self-control choices. More so, even following perseverance, children of giving up parents choose the no willpower option with probability $1 - p_2$ in the second period, even if they are of strong type. Finally, in the most extreme case when the distribution of types does not have a large mass at the top, and/or heritability is large, so that even condition (4.3) does not hold, children of giving up parents do not even attempt willpower in the first period. Once again, this includes the strong types.

4.2 Multiple Generations

So far, we have established that, conditional on type, children of giving up parents may exercise less self-control than children of persevering parents in the two-generation case. We will use this result to extend the two generations case to an infinite number of overlapping generations to find the limiting distribution of the self-control choices in the population. With multiple generations, the question of exactly what constitutes children's initial beliefs becomes important. On the hand, children could have 'short memory' and observe only their parents' behaviour but know nothing about their grandparents. On the other, children could learn their entire family history and dynasties based on beliefs could form. We start off by considering the former (as captured by Assumption 3).

For some generation $k \in \{2, 3, ...\}$, consider how agents' initial beliefs are formed. As in the two-generation case, if children observed their parents give up, they know with certainty their parents were not a strong type and therefore, their initial beliefs will be:

$$\rho_1^{k|G} = \Pr(\beta_{ki} \geq \beta^* \mid A_{t=1,\tau=2}^{(k-1)i} = G) = (1-\theta)(1-F(\beta^*)) = \rho_1^{2|G} \equiv \rho_1^G \tag{4.4}$$

Observing one's parents give up is a sufficient statistic of a sort. From the child's perspective, it does not matter what their grandfather did as they can be certain their father was not a strong type. Thus, starting with generation two, the initial beliefs of children of giving-up parents are constant across generations.

The inference of children of persevering parents is somewhat more involved. Under Assumption 3, children of persevering parents can no longer make the same inference as their parent's second period self as they do not know whether their father himself came from a persevering or giving-up family. However, in addition to observing their parent's choice, children also know the strategy played by the population, and can therefore still form a prior on their parent being a strong type. Denote by $\lambda_j(S_j)$ the proportion of generation j that perseveres at $t = 1$, $\tau = 2$ when the strategy $S = \{S_j\}_{j=1}^{\infty}$ is played by the population. Then, the initial beliefs of children of persevering parents in generation k will be:

$$\rho_1^{k|P} = \Pr(\beta_{ki} \geq \beta^* \mid A_{t=1,\tau=2}^{(k-1)i} = P) = \theta \tfrac{1-F(\beta^*)}{\lambda_{k-1}(S_{k-1})} + (1-\theta)(1-F(\beta^*)) \tag{4.5}$$

We turn now to determining $\lambda_j(S_j)$.

Proposition 3.5 *Under Assumptions* 3 *and* 4, *monotonicity of beliefs and if condition (4.3) holds, then, from generation* 2 *onwards, agents play according to the strategy profile* $\{S^P, S^G\} = \{S_2^P, S_2^G\}$.

Further, in this equilibrium, if

$$\tfrac{(1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} < \rho_2^* \tag{4.6}$$

then the fraction of generation k *persevering is:*

$$\lambda_k = (1-F(\hat{\beta}))[\theta + (1-\theta)(F(\beta'') - F(\hat{\beta}))]^{k-1} + \tfrac{1-F(\beta'')}{1-(F(\beta'')-F(\hat{\beta}))} \tag{4.7}$$

and in the limit:

$$\lim_{k\to\infty} \lambda_k = \tfrac{1-F(\beta'')}{1-(F(\beta'')-F(\hat{\beta}))}$$

Proof. Suppose the strategy profile $\{S^P, S^G\} = \{S_2^P, S_2^G\}$ is played in every generation $k \in \{2, 3, ...\}$. Suppose further that condition (4.3) holds and all agents attempt willpower in the first period, and assume for the moment that beliefs are such that the relevant informativeness constraints are satisfied for all generations $k \in \{2, 3, ...\}$. Then, by construction, the strategy profile $\{S^P, S^G\}$ is sequentially rational as it is not possible for any agent to do better than play according to $\{S^P, S^G\}$, provided their beliefs are consistent with $\{S^P, S^G\}$. This is the result of the fact that the strong types and the weak types play their dominant strategies and the intermediate types already persevere and choose willpower as frequently as possible in $\{S^P, S^G\}$. Therefore, to show that $\{S^P, S^G\}$ constitutes a Perfect Bayesian Equilibrium of the infinite overlapping generations game, we need to show that beliefs are consistent with $\{S^P, S^G\}$ for all $k \in \{2, 3, ...\}$.

As already established in equation (4.4), the initial beliefs of children of giving-up parents are the same across all generations, starting with generation 2. Therefore, in any generation k, children of giving up parents cannot do better than play according to $S_2^G \equiv S^G$. On the other hand, from equation (4.5), the initial beliefs of children of persevering parents depend on the proportion of parents' generation persevering. In generation k :

$$\rho_1^{k|P} = \Pr(\beta_{ki} \geq \beta^* \mid A_{t=1,\tau=2}^{(k-1)i} = P) = \theta \frac{1-F(\beta^*)}{\lambda_{k-1}(S^P, S^G)} + (1-\theta)(1-F(\beta^*)) \qquad (4.8)$$

Suppose such children play according to S^P and therefore persevere at $t = 1$, $\tau = 2$ if $\beta_{ki} \geq \hat{\beta}$. Then their second period beliefs, conditional on perseverance, would be:

$$\Pr(\beta_{ki} \geq \beta^* \mid A_{t=1,\tau=2}^{ki} = P,\ A_{t=1,\tau=2}^{(k-1)i} = P) = \frac{\theta \frac{1-F(\beta^*)}{\lambda_{k-1}(S^P, S^G)} + (1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \qquad (4.9)$$

For this to be an equilibrium, the updated second period beliefs in (4.9) must satisfy the informativeness constraint, i.e. we must have $\frac{\theta \frac{1-F(\beta^*)}{\lambda_{k-1}(S)} + (1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \geq \rho_2^*$. Recall that we assumed that the updated second period beliefs of the first generation satisfied the informativeness constraint, $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} \geq \rho_2^*$. Then, the updated second period beliefs of generation 2, conditional on perseverance, necessarily also satisfied the informativeness constraint $\frac{\theta \frac{1-F(\beta^*)}{1-F(\hat{\beta})} + (1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \geq \rho_2^*$, since $\theta \frac{1-F(\beta^*)}{1-F(\hat{\beta})} + (1-\theta)(1 - F(\beta^*)) \geq 1 - F(\beta^*)$. In fact, it is clear that the informativeness constraint is satisfied for all generations k, since for any $\lambda_k \leq 1$, $\theta \frac{1-F(\beta^*)}{\lambda_{k-1}(S)} + (1-\theta)(1 - F(\beta^*)) \geq 1 - F(\beta^*)$, and therefore, $\frac{\theta \frac{1-F(\beta^*)}{\lambda_{k-1}(S)} + (1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \geq \rho_2^*$, provided $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} \geq \rho_2^*$. Moreover, $\lambda_k \leq 1 - F(\hat{\beta})\ \forall k$ as no agent with $\beta < \hat{\beta}$ would ever persevere, no matter how high their beliefs might be.

Next, we determine λ_k. If the strategy profile $\{S^P, S^G\}$ is played, then the proportion of people persevering in any generation k consists of children of giving-up parents with $\beta_{ki} \geq \beta''$ and children of persevering parents with $\beta_{ki} \geq \hat{\beta}$:

$$\begin{aligned}
\lambda_k &= \lambda_{k-1} \Pr(\beta_{ki} \geq \hat{\beta} \mid A^{(k-1)i}_{t=1,\tau=2} = P) + (1-\lambda_{k-1}) \Pr(\beta_{ki} \geq \beta'' \mid A^{(k-1)i}_{t=1,\tau=2} = G) \qquad (4.10) \\
&= \lambda_{k-1}[\theta + (1-\theta)(1-F(\hat{\beta}))] + (1-\lambda_{k-1})(1-\theta)(1-F(\beta'')) \\
&= \lambda_{k-1}[\theta + (1-\theta)(F(\beta'') - F(\hat{\beta}))] + (1-\theta)(1-F(\beta''))
\end{aligned}$$

Equation (4.10) is a simple first order linear difference equation; its general solution is:

$$\lambda_k = M[\theta + (1-\theta)(F(\beta'') - F(\hat{\beta}))]^k + \frac{1-F(\beta'')}{1-(F(\beta'')-F(\hat{\beta}))} \qquad (4.11)$$

With $\lambda_1 = (1-F(\hat{\beta}))$, we can find the constant M in (4.11):

$$\begin{aligned}
M &= \frac{(1-F(\hat{\beta}))(1-\left[F(\beta'')-F(\hat{\beta})\right])-(1-F(\beta''))}{[1-(F(\beta'')-F(\hat{\beta}))][\theta+(1-\theta)(1-F(\hat{\beta}))]} \qquad (4.12) \\
&= \frac{1-F(\beta'')+F(\hat{\beta})\left[F(\beta'')-F(\hat{\beta})\right]}{[1-(F(\beta'')-F(\hat{\beta}))][\theta+(1-\theta)(1-F(\hat{\beta}))]} = \frac{1-F(\hat{\beta})}{\theta+(1-\theta)(1-F(\hat{\beta}))}
\end{aligned}$$

Substituting back into (4.11), we have:

$$\lambda_k = (1-F(\hat{\beta}))[\theta + (1-\theta)(F(\beta'') - F(\hat{\beta}))]^{k-1} + \frac{1-F(\beta'')}{1-(F(\beta'')-F(\hat{\beta}))} \qquad (4.13)$$

Since $(F(\beta'') - F(\hat{\beta})) < 1$ and therefore $[\theta + (1-\theta)(F(\beta'') - F(\hat{\beta}))] < 1$, taking the limit to infinity of (4.13), we have:

$$\begin{aligned}
\lim_{k\to\infty} \lambda_k &= (1-F(\hat{\beta})) \lim_{k\to\infty} [\theta + (1-\theta)(F(\beta'') - F(\hat{\beta}))]^{k-1} + \frac{1-F(\beta'')}{1-(F(\beta'')-F(\hat{\beta}))} \qquad (4.14) \\
&= \frac{1-F(\beta'')}{1-\left[F(\beta'')-F(\hat{\beta})\right]}
\end{aligned}$$

It is trivial to see that $\frac{1-F(\beta'')}{1-\left[F(\beta'')-F(\hat{\beta})\right]} < 1-F(\hat{\beta})$ as long as $F(\beta'') > F(\hat{\beta})$. ■

Proposition 3.5 states that from generation two onwards, all agents play the same strategy profile, $\{S^P, S^G\}$. The intuition is that children of giving-up parents have the same initial beliefs regardless of which generation they occur in, and so cannot do better than play S^G_2. Children of persevering parents have weakly increasing initial beliefs across generations, but since all intermediate (and strong) children of persevering parents in generation 2 already persevere and play a pure strategy in the second period, they cannot improve on the second generation strategy either.

As in the two generation case, in the multiple generations equilibrium children of persevering parents exercise more self-control than children of giving-up parents. There remains a fraction of the population - children of giving up parents with $\hat{\beta} \leq \beta_i < \beta''$ who do not persevere in equilibrium because their self-confidence is lowered by watching their parents give up. In the long run, the fraction of the population persevering settles to a steady state, $\frac{1-F(\beta'')}{1-\left[F(\beta'')-F(\hat{\beta})\right]}$, which is lower than the proportion of all agents who would have an incentive to persevere when their beliefs are not conditioned by parents' behaviour, $\frac{1-F(\beta'')}{1-\left[F(\beta'')-F(\hat{\beta})\right]} < 1 - F(\hat{\beta})$.

In Proposition 3.5 we assumed that children have short memory and observe or recall only the actions of their immediate parent, and not of any further ancestors. However, it is clear from the proof of this proposition that even if children could recall the entire family history, the result would be the same. The only reason family history could matter is if a history of perseverance improved the child's self-confidence and if this higher self-confidence translated into more self-control. However, children of giving up parents know with certainty that their parents were not of strong type, so that knowing the behaviour of the grandfathers has no effect on the initial beliefs of these children. For children of persevering parents, their initial beliefs do in fact increase with the length of the spell of perseverance in the family history, but this does not result in more self-control simply because even children with only the immediate parent (and not parent's parent) persevering have sufficient self-confidence to persevere in the first period for all $\beta_i \geq \hat{\beta}$ and choose willpower with probability one in the second period, conditional on observing perseverance. This is the best possible outcome so even agents with higher initial beliefs cannot do better than that. This results from the fact that self-control choice is discrete and only an increasing step function of initial beliefs. If the amount of self-control exercised were a continuous choice, family history, and particularly the length of spells of perseverance, would matter, all else equal.

5 Overlapping Generations Model with Parental Altruism

In this section we extend our basic set-up to account for a form of paternalistic altruism, or intergenerational empathy, in which children's welfare may enter directly into parents' utility. Then, in addition to parents' choices having an effect on children's choices, through their initial beliefs, expected children's behaviour may have an effect on parents' behaviour. In particular, if giving up on willpower reveals the parent as being a weak type and consequently lowers his offspring's initial self-confidence, a parent who cares about his child's wellbeing may have an additional incentive to persevere.

5.1 Two Generations with Parental Altruism

We start off again by considering the case of two generations only: a first generation of parents followed by a second generation of children. We continue to assume that children observe period one actions of their parents. We now also assume that the anticipated children's utility enters directly into parents' payoffs at $t = 1$, $\tau = 2$. Denote by $E[u_{2i}(S^{2i}) \mid \beta_{1i}, A^{1i}_{t=1,\tau=2}]$ the parent's expectation of his child's utility when the parent chooses action A^{1i} at $t = 1$, $\tau = 2$ and the child plays the strategy S^{2i}; this expectation is taken at $t = 1$, $\tau = 2$, when the parent's type is momentarily revealed to the parent, so it is conditional on parent's type, β_{1i}. Then, for a parent of type β_{1i}, the period 1 payoff to G becomes $b + \mu E[u_{2i}(S^{2i}) \mid \beta_{1i}, G]$ and to P, $B - c/\beta_{1i} + \mu E[u_{2i}(S^{2i}) \mid \beta_{1i}, P]$, where $\mu \in [0,1]$ is a discount factor which accounts for imperfect empathy between parents and children[34].

When children observe the first period choices of their parents, it is the parents' choice between P and G that determines the child's initial beliefs. We therefore think it reasonable that the anticipated child welfare should enter parents' utility at the time of the relevant choice, i.e. at $t = 1$, $\tau = 2$. We refer to this scenario as anticipatory empathy. One might argue that it is the actual child's utility that should enter into parents' utility when children make their choices, but insofar as parents make their first period choices before children start to make any decisions, it is still the children's expected welfare that will affect the parents' choice between P and G.

Proposition 3.6 *Under Assumption* 4, *monotonic beliefs and if the distribution of types, $F(\beta)$, satisfies*

$$
\begin{aligned}
&(1 - F(\beta^*))(B - c + \delta(B - c)) + (F(\beta^*) - F(\bar{\beta}))(B - c + \delta b) + F(\bar{\beta})(b + \delta a) \qquad (5.1)\\
&\qquad + \mu(1 - \theta) \times \\
&\left[\begin{array}{c} (1 - F(\bar{\beta})) \left[(1 - F(\beta^*))(B - c + \delta(B - c)) + (F(\beta^*) - F(\hat{\beta}))(B - c + \delta b) + F(\hat{\beta})(b + \delta a) \right] \\ + F(\bar{\beta}) \left[\begin{array}{c} (1 - F(\beta^*))(B - c + \delta(p_2(B - c) + (1 - p_2)a)) \\ + (F(\beta^*) - F(\beta''))(B - c + \delta(p_2 b + (1 - p_2)a)) + F(\beta'')(b + \delta a) \end{array} \right] \end{array} \right] \\
&+ \mu\theta \left[(1 - F(\beta^*))(B - c + \delta(B - c)) + (F(\beta^*) - F(\hat{\beta}))(B - c + \delta b) + F(\hat{\beta})(b + \delta a) \right] \\
&\qquad \geq \tfrac{a}{\gamma} + \delta a + \mu(1 + \delta)a
\end{aligned}
$$

[34] By assuming $\mu \in [0,1]$, we ignore the possibility that parents care more about their children than themselves.

and

$$\frac{1-F(\beta^*)}{1-F(\bar{\beta})} \geq \rho_2^* \tag{5.2}$$

where

$$\bar{\beta} = c\{B - b + \delta(b-a) + \mu(1-\theta)[\delta(1-p_2)[(1-F(\beta^*))(B-c-a) + (F(\beta^*) - F(\beta''))(b-a)] + (F(\beta'') - F(\hat{\beta}))(B-c-b+\delta(b-a))]\}^{-1} \tag{5.3}$$

a semi-separating equilibrium exists. In this equilibrium, parents play the strategy S_1^e*, according to which they:*

- *At* $t = 1,\ \tau = 1$ *: Attempt willpower.*
- *At* $t = 1,\ \tau = 2$ *: Persevere if* $\beta_{1i} \geq \bar{\beta}$*, where* $\bar{\beta} < \hat{\beta}$*. Give up otherwise.*
- *At* $t = 2,\ \tau = 1$ *: Choose willpower with probability* 1 *if perseverance was observed in the previous period; otherwise, choose no willpower.*
- *At* $t = 2,\ \tau = 2$ *: Persevere if* $\beta_{1i} \geq \beta^*$*. Give up otherwise.*

Children's generation play according the strategy profile $\{S_2^P, S_2^G\}$*, which is the same as in Proposition (3.4).*

Proof of Proposition 3.6 can be found in Appendix B.

In Proposition 3.6, we show that in the equilibrium of a two-generation game, parental altruism raises the self-control outcomes of the parents' generation, with an additional fraction of parents, $F(\hat{\beta}) - F(\bar{\beta})$, persevering in the first period and therefore attempting willpower in the second. The increased incentive for self-control comes from the fact that parental perseverance raises children's initial self-confidence, which improves the self-control outcomes in the children's generation and therefore raises parents' expectation of children's utility. Since this game ends with generation two, children's payoffs are not affected by altruism and they continue to play the same strategy profile $\{S_2^P, S_2^G\}$ as they did before altruism was introduced. The only difference is that the initial beliefs of children of altruistic parents are somewhat lower than of non-altruistic parents. This is due to the fact that altruism is common knowledge and children of persevering parents rationally deduce that some weaker parents would have persevered precisely in order to induce higher beliefs among their children. Still, since we continue to assume that the parents' informativeness constraint is satisfied sufficiently for them to play a pure strategy following perseverance (condition (5.2) holds), the informativeness constraint will also necessarily be satisfied for children of persevering parents, who would also play a pure strategy. Equilibrium would still be possible if condition (5.2) did not hold, and children would continue to play the same strategy, although they would have slightly higher initial beliefs. However, in the case of altruism, the assumption that parents play a pure strategy becomes non-trivial in the sense that it affects the direction in which the introduction of children's utility affects parental self-control outcomes. As we will show in the next section, introduction of parental altruism can actually reduce the amount of self-control exercised by the parents if parents play a mixed strategy.

For parents to even reach the second information set, condition (5.1) ensures that the distribution of types and therefore the initial beliefs of the first generation, are sufficiently high for parents to attempt willpower. Otherwise a pooling equilibrium would be played in which no willpower was chosen in both generations. In the case that parents attempt willpower but children of giving up parents have insufficient self-confidence to try willpower themselves, a pure semi-separating equilibrium is possible. In this equilibrium an even larger fraction of parents persevere, since giving up has a more punitive effect on the expected self-control outcomes, and therefore utility, of children.

5.2 Multiple Generations with Parental Altruism

Suppose now there is an infinite number of overlapping generations, so that all children are also parents who have their own children. With infinite generations, even if it is only the utility of the immediate offspring that enters directly into parent's utility, the children will also care about their own children, and so on ad infinitum. Therefore, any parent will by default care about the discounted utilities of all of his future descendants. Suppose all generations $k = \{1, 2, ...\}$ play according to the strategy profile $\{S\}_{k=1}^{\infty}$. Denote by $V_{k+1,i}(\beta_{ki}, A^{ki}_{t=1,\tau=2}, S^{(k+1)i})$ the expected utility of all future generations in dynasty i, starting in generation $k+1$, from the perspective of agent k in dynasty i, who is of type β_{ki} and chooses action A^{ki} at $t = 1$, $\tau = 2$.

Proposition 3.7 *Under Assumption* (4), *monotonic beliefs and if the distribution of types, $F(\beta)$, satisfies*

$$\begin{aligned}(1-\theta)(1-F(\beta^*))(B-c+\delta p_2^e(B-c-a)+\mu V(\beta \geq \beta^*, P)) \\ +(1-\theta)(F(\beta^*)-F(\beta''))(B-c+\delta p_2^e(b-a)+\mu V(\beta'' \leq \beta < \beta^*, P)) \\ +(\theta+(1-\theta)F(\beta''))(b+\mu V(\beta < \beta'', G)) \geq a/\gamma + a/(1-\delta)\end{aligned} \tag{5.4}$$

and

$$\frac{1-F(\beta^*)}{1-F(\bar{\beta})} \geq \rho_2^* \tag{5.5}$$

a stationary semi-separating equilibrium exists. In this equilibrium, from generation two onwards, all generations play the strategy profile $\{S^{eP}, S^{eG}\}$. In any generation $k \in \{2, 3, ..\}$, children of persevering parents play the following strategy S^{eP}:

- *At $t = 1$, $\tau = 1$: Attempt willpower.*
- *At $t = 1$, $\tau = 2$: Persevere if $\beta_{ki} \geq \tilde{\beta}$. Give up otherwise.*
- *At $t = 2$, $\tau = 1$: Choose willpower with probability 1 if perseverance was observed in the previous period; otherwise, choose no willpower.*
- *At $t = 2$, $\tau = 2$: Persevere if $\beta_{ki} \geq \beta^*$. Give up otherwise.*

Children of giving-up parents play the following strategy S^{eG} :

- *At $t = 1$, $\tau = 1$: Attempt willpower.*
- *At $t = 1$, $\tau = 2$: Persevere if $\beta_{ki} \geq \tilde{\beta}$ and $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\tilde{\beta})} \geq \rho_2^*$, or if $\beta_{2i} \geq \beta''$ and if $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\tilde{\beta})} < \rho_2^*$ but $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\beta'')} = \rho_2^*$ for some $\beta'' \in (\tilde{\beta}, \beta^*)$. Give up otherwise.*
- *At $t = 2$, $\tau = 1$: Choose willpower with probability 1 if perseverance was observed in the previous period and $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\tilde{\beta})} \geq \rho_2^*$; Choose willpower with probability p_2^e if willpower was observed in previous period and $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\beta'')} = \rho_2^*$. Otherwise, choose no willpower.*
- *At $t = 2$, $\tau = 2$: Persevere if $\beta_{ki} \geq \beta^*$. Give up otherwise.*

Where,

$$\tilde{\beta} = c\left[\frac{B-b-c/\beta''+\delta(b-a)}{1-\mu D} + \frac{c}{\beta''}\left(1 + (1-\theta)(F(\beta'') - F(\tilde{\beta}))\frac{1-\beta''}{(1-\mu D)}\right) + \frac{\theta(1-\beta'')}{(1-\mu\theta-\mu(1-\theta)(F(\beta'')-F(\tilde{\beta})))}\right]^{-1} \tag{5.6}$$

and

$$p_2^e = 1 - \frac{B-b-c/\beta''+\delta(b-a)+\mu(1-\theta)(F(\beta'')-F(\tilde{\beta}))c\frac{1-\beta''}{\beta''}}{\delta(b-a)(1-\mu D)} \tag{5.7}$$

where $D \equiv \theta + (1-\theta)(F(\beta^*) - F(\tilde{\beta})) + (1-\theta)(1-F(\beta^*))\frac{B-c-a}{b-a}$.

Generation 1 play the same strategy as children of persevering parents, S^{eP}.

Proof of Proposition 3.7 can be found in Appendix B.

In Proposition 3.7, we describe the stationary semi-separating equilibrium of the infinite generations game with parental altruism. We show that parental altruism raises the fraction of population that perseveres in the first period by an additional

$F(\hat{\beta}) - F(\tilde{\beta})$ in every generation. More so, a larger fraction of children of persevering parents themselves persevere in the first period than either in the two-generation or no-altruism case, since parental actions can now have an effect on all future generations[35], i.e. $\tilde{\beta} \leq \bar{\beta} \leq \hat{\beta}$. The fraction of children of giving-up parents who persevere in this equilibrium, $1 - F(\beta'')$, is the same as in the two-generation and no-altruism case, as it is driven by the same informativeness constraint. As a result, we also find that parental altruism raises the faction of the population that perseveres in the long run: $\lim_{k\to\infty} \lambda_k^e = \frac{1-F(\beta'')}{1-(F(\beta'')-F(\tilde{\beta}))} > \lim_{k\to\infty} \lambda_k$.

Although the fraction of children of giving up parents who persevere is unchanged, the probability with which these agents then choose willpower in the second period is actually reduced, $p_2^e < p_2$. This somewhat counterintuitive outcome is the result of the fact that children of giving-up parents play a mixed strategy: because the anticipated children's utility enters into parent's utility at $t = 1, \tau = 2$ and is greater for the children of persevering parents than of giving up, for the β'' agent to remain indifferent between P and G, he must select willpower following P with a lower probability in the second period. In other words, the additional utility gained from children's welfare following parental perseverance, reduces the value of reputational stake needed to compensate agent β'' for costly perseverance in the first period. Such parents substitute future children's welfare for utility gained form exercise of additional self-control in later stage of their life.

5.3 Late Parenthood

When children's welfare enters into parents' utility, the timing of the overlap between parents' and children's generations, and therefore which actions children observe and when parents 'experience' children's utility, becomes potentially important. We therefore investigate an alternative assumption on the timing of the overlap: rather than children observing period one actions of their parents, we now assume that children are born one period later so that children's period zero, in which they only observe, coincides with parents' period two. The question is whether children's monitoring of the second half of the parents' life can induce the parents to exercise more self-control throughout their lifetime.

For simplicity, we return to the two generation case. Denote by $E[u_{2i}^{LP}(S^{2i}) \mid \beta_{1i}, A^{1i}_{t=2,\tau=2}]$ the expectation of β_{1i} parent of his child's utility given that the parent chooses action $A^{1i}_{t=2,\tau=2}$ at $t = 2, \tau = 2$ and the child plays according to strategy S^{2i}.

[35] Since $V(\beta, P) \geq V(\beta, G)$ in all generations since parental perseverance weakly increases the probability of child perseverance, it must be that $V(\beta, P) - V(\beta, G) \geq E[u_{2i}(P, S^{2P}) \mid \bar{\beta}] - E[u_{2i}(G, S^{2G}) \mid \bar{\beta}]$ fo all $\theta \neq 0$.

Proposition 3.8 *Under Assumption* (4), *monotonic beliefs and if the distribution of types,* $F(\beta)$, *satisfies*

$$
\begin{aligned}
&(1-F(\beta^{**}))(B-c+\delta(B-c+\mu E[u_{2i}^{LP}(S^{2P}) \mid \beta_{1i} \geq \beta^{**}, P])) \\
&+(F(\beta^{**})-F(\hat{\beta}))(B-c+\delta(b+\mu E[u_{2i}^{LP}(S^{2G}) \mid \hat{\beta} \leq \beta_{1i} < \beta^{**}, G])) \\
&+F(\hat{\beta})(b+\delta a+\mu E[u_{2i}^{LP}(S^{NW}) \mid \beta_{1i.} < \hat{\beta}, NW)) \geq \tfrac{a}{\gamma}+\delta a+(1+\delta)a
\end{aligned}
\tag{5.8}
$$

and

$$\frac{1-F(\beta^{*})}{1-\min\{F(\hat{\beta}),F(\beta^{**})\}} \geq \rho_2^{*} \tag{5.9}$$

a semi-separating equilibrium exists. In the equilibrium, parents will play according to the strategy S_1^{LP} :

- *At* $t=1$, $\tau=1$: *Attempt willpower.*
- *At* $t=1$, $\tau=2$: *Persevere if* $\beta_{1i} \geq \min\{\hat{\beta}, \beta^{**}\}$. *Give up otherwise.*
- *At* $t=2$, $\tau=1$: *Choose willpower with probability* 1 *if perseverance was observed in the previous period; otherwise, choose no willpower.*
- *At* $t=2$, $\tau=2$: *Persevere if* $\beta_{1i} \geq \beta^{**}$, *where* $\beta^{**} < \beta^{*}$. *Give up otherwise.*

Children will play according to the strategy profile $\{S_2^{NW}, S_2^{G}, S_2^{P}\}$ *where* S_2^{G} *and* S_2^{P} *are as before, and* S_2^{NW} *and* S_2^{G} *are equivalent.*

Proof of Proposition 3.8 can be found in Appendix B.

Proposition 3.8 states that when children observe the second period actions of their parents, all parents with $\beta_{1i} \geq \beta^{**}$, where $\beta^{**} < \beta^{*}$, persevere at the end of the second period. Having children observe the second period provides the (intergenerational) reputational effect that was otherwise absent at the end of the game. Thus, in period one, intermediate types persevere in order to induce favourable beliefs in their own future selves, and in period two, some intermediate types persevere in order to induce higher beliefs in their children. Whether under successive generations children have an effect on parents' period one behaviour is ambiguous. If $\beta^{**} > \hat{\beta}$, then only those parents with $\beta_{1i} \geq \hat{\beta}$ persevere in the first period, which is the same threshold level as in the no altruism case. If $\beta^{**} < \hat{\beta}$, then all parents with $\beta_{1i} \geq \beta^{**}$ persevere in both periods. This outcome can arise when the difference between expected utilities of children of persevering and giving up parents is larger than the maximum reputational stake that can be attained by agent in the absence of effects on children. This is possible when the differences in payoffs to persevering and giving up is considerably larger than between giving up and not attempting willpower, and when the intergenerational discount factor μ is high, and it implies that even some of the weak types who would never persevere without altruism, persevere in both periods.

Although we continue to focus on the equilibria in which parents play a pure strategy, which is the reason we always assume that $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} \geq \rho_2^*$, where $\hat{\beta}$ is the last parent with an incentive to persevere in the first period, the mixed version of the parents' strategy is worth a mention in this case. As we showed in the proof of Proposition 3.8, the introduction of discounted child welfare into parents' utility in the second period has the effect of lowering the minimum level of self-confidence required to attempt willpower in the second period, $\rho_2^{**} < \rho_2^*$. In a mixed equilibrium, this can potentially allow more parents to persevere in the first period whilst keeping the second period self-confidence at its new threshold level, ρ_2^{**}, which in turn increases the probability with which the second period selves attempt willpower. In the last subperiod, all parents with $\beta_{1i} \geq \beta^{**}$ persevere. Thus, if parents originally play a mixed strategy, the introduction of children's monitoring of parental actions in period two increases the amount of self-control exercised by the parents at every information set.

6 Conclusion

In this chapter, we developed an overlapping generations version of the Benabou and Tirole (2004) self-signaling model of personal rules to see how parental behaviour can affect self-control outcomes of their children. We find that children of more self-controlled parents have higher initial self-confidence, which conditional on type, leads them to exercise more self-control themselves. We find the fraction of the population who do not persevere in the long run equilibrium simply because their initial self-confidence is lowered by observation of poor parental choices. Further, introduction of parental altruism increases the amount of self-control exercised by the parents, especially so if children observe parents' behaviour in the later stages of parents' lives.

A potential way to advance this chapter would be to investigate different inheritance mechanisms for the discount rates between parents and children. In particular, our inheritance mechanism leaves the distribution of types the same across generations, but it may be curious to see if an alternative transmission mechanism could lead to evolution of distribution of types over generations, and ultimately a steady state independent of the initial distribution of types.

7 Appendix A

7.1 *Proof of Proposition 3.2*

Proof. The proof is analogous to that of Proposition 1 in Benabou and Tirole. Suppose the strategy profile S^* is played, and consider the agent's choice in period 2. At the end of the second period, $t = 2$, $\tau = 2$, only the strong types, with $\beta_i \geq \beta^*$, persevere. Then, since $b < \frac{a}{\gamma} < B - c$, an agent will only attempt willpower at the beginning of the second period, $t = 2$, $\tau = 1$, if he is sufficiently confident of being a strong type and persevering at the end of the period, i.e. $\rho_2^i = \Pr(\beta_i \geq \beta^* \mid A_{t=1}^i) \geq \rho_2^* = \frac{a/\gamma - b}{B-c-b}$, where ρ_2^* solves $\rho_2^*(B-c) + (1-\rho_2^*)b = \dfrac{a}{\gamma}$, as before.

Suppose agents' beliefs at the beginning of period 2 are in fact: $\Pr(\beta_i \geq \beta^* \mid A_{t=1}^i) = \frac{1-F(\beta^*)}{1-F(\hat{\beta})}$ and $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} \geq \rho_2^*$ by condition (3.2), i.e. the informativeness constraint is satisfied. Then, consider the agent's choice at $t = 1$, $\tau = 2$. With monotonic beliefs, it is again the dominant strategy for the strong types to persevere. If , according to S^*, willpower is chosen with probability one at $t = 2$, $\tau = 1$ if P was observed in the first period, and with zero probability otherwise, all agents with $\beta_i \in [\hat{\beta}, \beta^*)$, where $\hat{\beta} = \frac{c}{B-b+\delta(b-a)}$, will also have an incentive to persevere as $B - \frac{c}{\beta_i} + \delta b \geq b + \delta a$ for all $\beta_i \geq \hat{\beta}$.

Therefore, the strategy profile S^* is optimal given beliefs are high enough to satisfy the informativeness constraint. It remains to check that beliefs are correct in equilibrium.

If, according to S^*, all agents with $\beta_i \geq \hat{\beta}$ persevered at $t = 1$, $\tau = 2$, then the updated beliefs at $t = 2$, $\tau = 1$ following perseverance would be:

$$\rho_2^P = \Pr(\beta_i \geq \beta^* \mid A_{t=1}^i = P) = \frac{1-F(\beta^*)}{1-F(\hat{\beta})} \tag{7.1}$$

And following giving up:

$$\rho_2^G = \Pr(\beta_i \geq \beta^* \mid A_{t=1}^i = G) = 0 \tag{7.2}$$

Condition (3.2) ensures that equilibrium second period beliefs satisfy the informativeness constraint. Otherwise, the second period self-confidence following perseverance would not be high enough to induce an agent to choose willpower, which cannot be an equilibrium for the intermediate types as they would only incur a cost, $b - (B - \frac{c}{\beta i})$, from persevering in the first period, if there is a prospect of a gain, $\delta(b-a)$, in the second period.

Finally, at the beginning of period 1, an agent would choose to attempt willpower if:

$$(1-F(\beta^*))(B-c+\delta(B-c)) + (F(\beta^*)-F(\hat{\beta}))(B-c+\delta b) + F(\hat{\beta})(b+\delta a) \geq \tfrac{a}{\gamma} + \delta a \tag{7.3}$$

The LHS of (7.3) is the ex ante expected payoff to willpower for any agent if S^* is played in equilibrium. The RHS is the overall payoff to no willpower, with the immediate payoff overweighted by γ. If initial beliefs are such that agents are sufficiently confident of persevering should they attempt willpower, i.e. (7.3) holds, then willpower is initially attempted by all agents. Otherwise, no-one attempts willpower and there is a pooling equilibrium on NW.

Given that in the semi-separating equilibrium, S^* is played by the population and beliefs are updated according to S^*, no agent j can profitably deviate. The strong types have a dominant strategy and will always persevere, regardless of initial or posterior beliefs, and regardless of other types' actions. The weak types also have a dominant strategy and will never persevere, as even the prospect of willpower with certainty in period 2 is not enough to overcome their low discount rates in period 1. All the intermediate types already persevere in the proposed equilibrium, and since perseverance yields greater payoff when followed by willpower than giving up, it cannot be profitable for the intermediate types to deviate by giving up, nor by choosing no willpower following perseverance.

Therefore, since the actions of all types according to S^* are sequentially rational and beliefs are correct in equilibrium given S^*, we have a Perfect Bayesian Equilibrium. ■

7.2 *Proof of Proposition 3.3*

Proof. The proof is analogous to that of Proposition 3.2 with the following amendment. Suppose the strategy profile S^{m*} is played and at $t = 2$, $\tau = 1$, agents play willpower with probability p_2 conditional on observing perseverance in the first period, and with zero probability otehrwise Then, at $t = 1$, $\tau = 2$, for any $\beta_i < \beta^*$, persevering yields $B - c/\beta_i + \delta(p_2 b + (1-p_2)a)$, and giving up yields $b + \delta a$. We need to determine the last agent to persevere in the first period, call him β', and the probability with which willpower is chosen in the second period, p_2.

First, in order for an agent to randomise at $t = 2$, $\tau = 1$, he must be indifferent between willpower and no willpower, i.e. his posterior self-confidence must be exactly ρ_2^*. Therefore, the last agent to persevere in the first period must be such β' that solves the informativeness constraint with equality:

$$\frac{1-F(\beta^*)}{1-F(\beta')} = \rho_2^* \tag{7.4}$$

Second, in order for such equilibrium to be deviation-proof, no agent with $\beta_i < \beta'$ must have an incentive to persevere. Therefore, at $t = 2$, $\tau = 1$, the agent must randomise with such probability p_2 which makes β' the last agent who has the incentive to persevere in the first period. Specifically, p_2 must solve: $B - c/\beta' + \delta(p_2 b + (1-p_2)a) = b + \delta a$. Therefore,

$$p_2 = \frac{b - B + c/\beta'}{\delta(b-a)} \tag{7.5}$$

Then, by construction, no agent with $\beta_i < \beta'$ has an incentive to deviate by persevering.

Can any agent do better by giving up? Recall that by construction, no agent below $\hat{\beta}$ has an incentive to persevere even if perseverance leads to willpower being chosen with probability one in period two. Since in the proposed mixed equilibrium, willpower is only played with probability p_2 in period two, the reputational stake is smaller than in the pure equilibrium, and therefore, it must be the case that $\beta' > \hat{\beta}$. Otherwise, all agents with $\beta' < \beta_i < \hat{\beta}$ would do better by giving up (also, p_2 would be negative).

Since in this equilibrium, the agent is indifferent between willpower and no willpower in period two, he cannot profitably deviate at $t = 2$, $\tau = 1$ either.

Finally, if S^{m*} is played, an agent would have an incentive to attempt willpower at $t = 1$, $\tau = 1$ if the expected payoff to willpower is greater than the expected payoff to no willpower:

$$(1 - F(\beta^*))(B - c + \delta(p_2(B - c) + (1 - p_2)a)) + (F(\beta^*) - F(\beta'))(B - c + \delta(p_2 b + (1 - p_2)a)) \tag{7.6}$$

$$+F(\beta')(b + \delta a) \geq \frac{a}{\gamma} + \delta a$$

■

7.3 *Proof of Proposition 3.4*

Proof. Suppose condition (4.3) holds and all agents attempt willpower in the first period. We already know from Proposition 3.2 that the strategy profile S_2^P is sequentially rational, provided the informativeness constraint is satisfied. From Proposition 3.3, we can easily deduce that the strategy S_2^G is also sequentially rational if the informativeness constraint is satisfied. Therefore, it remains to show that when the strategy profile $\{S_2^P, S_2^G\}$ is played, the informativeness constraint is satisfied for children of both persevering and giving up parents.

Consider the second period beliefs of the children's generation. Firstly, giving up in the first period means that the agent cannot be a strong type, regardless of his parents' actions, meaning that agent's second period beliefs would be:

$$\rho_2^G = \Pr(\beta_{2i} \geq \beta^* \mid A_{t=1,\tau=2}^{2i} = G) = 0 \tag{7.7}$$

Secondly, under S_2^P, children of persevering parents themselves persevere if $\beta_{2i} \geq \hat{\beta}$ so that their updated second period beliefs, conditional on observing P in the first period, would be:

$$\rho_2^{P,P} = \Pr(\beta_{2i} \geq \beta^* \mid A^{2i}_{t=1,\tau=2} = P,\ A^{1i}_{t=1,\tau=2} = P) = \frac{\theta\frac{1-F(\beta^*)}{1-F(\hat{\beta})} + (1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \tag{7.8}$$
$$= \frac{1-F(\beta^*)}{[1-F(\hat{\beta})]^2}(1-(1-\theta)F(\hat{\beta}))$$

Since we assumed that the informativeness constraint was satisfied in generation 1, $\frac{1-F(\beta^*)}{1-F(\hat{\beta})} \geq \rho_2^*$, the informativeness constraint will necessarily be satisfied for persevering children of persevering parents:

$$\frac{\theta\frac{1-F(\beta^*)}{1-F(\hat{\beta})} + (1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \geq \rho_2^* \tag{7.9}$$

Thirdly, for children of giving up parents, two scenarios can arise. Suppose children of giving-up parents play according to pure strategy equilibrium at $t = 1$, $\tau = 2$. Then, according to S_2^G, their updated second period beliefs following P in the first period would be:

$$\rho_2^{P,G} = \Pr(\beta_{2i} \geq \beta^* \mid A^{2i}_{t=1,\tau=2} = P,\ A^{1i}_{t=1,\tau=2} = G) = \frac{(1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \tag{7.10}$$

This can only be an equilibrium if the informativeness constraint holds:

$$\frac{(1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \geq \rho_2^* \tag{7.11}$$

If (7.8) does not hold, then we know from Proposition 3.3 that there cannot be a pure equilibrium (beliefs are not high enough for all intermediate types to persevere, and a strategy where some or none of the intermediate types persevere is not deviation proof when willpower is chosen with probability one in the second period). In that case, according to S_2^G, children of giving up parents play mixed: at $t = 1$, $\tau = 2$, they persevere for all $\beta_{2i} \geq \beta''$; at $t = 2$, $\tau = 1$, they choose willpower with probability $p_2 = \frac{b-B+c/\beta''}{\delta(b-a)}$, where β'' solves the informativeness constraint with equality:

$$\rho_2^{P,G} = \Pr(\beta_{2i} \geq \beta^* \mid A^{2i}_{t=1,\tau=2} = P,\ A^{1i}_{t=1,\tau=2} = G) = \frac{(1-\theta)(1-F(\beta^*))}{1-F(\beta'')} = \rho_2^* \tag{7.12}$$

Finally, consider the agent's choice at the beginning of the game, at $t = 1$, $\tau = 1$. It was optimal for generation 1 to choose willpower in the first period, then it is necessarily optimal for children of persevering parents to choose willpower as well, since they have better odds of being a strong type than the underlying distribution. If S_2^G is played in equilibrium, children of giving up parents would choose willpower in the first period if:

$$(1-\theta)(1-F(\beta^*))(B-c+\delta p_2(B-c)+(1-p_2)a)+ \quad (7.13)$$
$$(1-\theta)(F(\beta^*)-F(\beta''))(B-c+\delta(p_2b+(1-p_2)a))$$
$$+[\theta+(1-\theta)F(\beta'')](b+\delta a) \geq \tfrac{a}{\gamma}+\delta a$$

The LHS in condition (7.13) above is the expected payoff to choosing willpower in the first period for a child of giving up parents, given that the mixed strategy equilibrium is played in period two. The RHS is the payoff to no willpower.

If it is optimal for a child of giving-up parents to attempt willpower under mixed strategy at $t=2$, $\tau=1$, it will also be optimal to attempt willpower if pure strategy is expected. ■

8 Appendix B

8.1 *Proof of Proposition 3.6*

Proof. Suppose the strategy profile $\{S_1^e; S_2^P, S_2^G\}$ is played.

Step 1. Consider the parents' generation. The $t=2$ payoffs are unaffected by the introduction of empathy, so at $t=2$, $\tau=2$ only the strong types, with $\beta_{1i} \geq \beta^*$, persevere; and at $t=2$, $\tau=1$, agents attempt willpower only if their self-confidence is above the threshold ρ_2^*. Consider what happens at $t=1$, $\tau=2$. For the strong types, according to S_1^e, perseverance P yields $B-\frac{c}{\beta_{1i}}+\delta(B-c)+\mu E[u_{2i}(S^{2P}) \mid \beta_{1i} \geq \beta^*, P]$, and giving up G yields $b+\delta a+\mu E[u_{2i}(S^{2G}) \mid \beta_{1i} \geq \beta^*, G]$. The expectations over the child's utility, $E[u_{2i}(S^{2P}) \mid \beta_{1i} \geq \beta^*, P]$ and $E[u_{2i}(S^{2G}) \mid \beta_{1i} \geq \beta^*, G]$, are taken from the strong type parent's perspective at $t=1$, $\tau=2$, given that children play the strategy profile $\{S_2^P, S_2^G\}$.

As long as $E[u_{2i}(S^{2P}) \mid \beta_{1i} \geq \beta^*, P] \geq E[u_{2i}(S^{2G}) \mid \beta_{1i} \geq \beta^*, G]$, it remains the dominant strategy for the strong types to persevere.

For any $\beta_{1i} < \beta^*$, according to S_1^e, persevering yields $B-\frac{c}{\beta_{1i}}+\delta b+\mu E[u_{2i}(S^{2P}) \mid \beta_{1i} < \beta^*, P]$, and giving up yields $b+\delta a+\mu E[u_{2i}(S^{2G}) \mid \beta_{1i} < \beta^*, G]$. By continuity, there exists some $\bar{\beta}$ parent who is indifferent between P and G :

$$\bar{\beta} = \frac{c}{B-b+\delta(b-a)+\mu(E[u_{2i}(S^{2P}) \mid \bar{\beta}, P]-E[u_{2i}(S^{2G}) \mid \bar{\beta}, G])} \quad (8.1)$$

Since, by construction, $\bar{\beta}$ is the last parent who wishes to persevere under altruism conditional on willpower being chosen with certainty upon observing P in period two, no agent with $\beta_{1i} < \bar{\beta}$ would have the incentive to deviate from S_1^e by persevering. Similarly, since we assume that the informativeness constraint is satisfied in parents' generation, $\frac{1-F(\beta^*)}{1-F(\bar{\beta})} \geq \rho_2^*$, no agent with $\beta_{1i} \geq \bar{\beta}$ can do better by giving up or playing a mixed strategy, since for such agents, P yields a strictly higher payoff than G at $t=1$, $\tau=2$, and willpower is already chosen with probability one at $t=2$, $\tau=1$, conditional on observing perseverance, which attains the maximum reputational stake.

Then, to show that S_1^e is optimal, we need to show that $E[u_{2i}(S^{2P}) \mid \beta_{1i} > \bar{\beta}, P] \geq E[u_{2i}(S^{2G}) \mid \beta_{1i} > \bar{\beta}, G]$, which would also imply that $\bar{\beta} \leq \hat{\beta} = \dfrac{c}{B-b+\delta(b-a)}$.

Step 2. Consider the children's generation. For children of persevering parents, their initial beliefs are:

$$\rho_1^{2|P} = \Pr(\beta_{2i} \geq \beta^* \mid A^{1i}_{t=1,\tau=2} = P) = \theta\tfrac{1-F(\beta^*)}{1-F(\bar{\beta})} + (1-\theta)(1-F(\beta^*)) \tag{8.2}$$

And for children of giving up parents:

$$\rho_1^{2|G} = \Pr(\beta_{2i} \geq \beta^* \mid A^{1i}_{t=1,\tau=2} = G) = (1-\theta)(1-F(\beta^*)) \tag{8.3}$$

The initial beliefs of children of persevering parents are actually somewhat lower under empathy since children take into account that a larger proportion of parents persevere in order to induce 'good' beliefs in their children. Still, since children of persevering parents have higher initial beliefs than their parents, and the parents already play the best possible strategy given $\frac{1-F(\beta^*)}{1-F(\bar{\beta})} \geq \rho_2^*$, the children cannot do better than persevere for $\beta_{2i} \geq \hat{\beta}$, and play willpower with probability one in period two, conditional on observing perseverance, which is exactly what S_2^P prescribes. More so, if children of persevering parents play according to S_2^P, their updated second period beliefs must necessarily satisfy the informativeness constraint, since $\dfrac{\theta\frac{1-F(\beta^*)}{1-F(\bar{\beta})}+(1-\theta)(1-F(\beta^*))}{1-F(\hat{\beta})} \geq \rho_2^*$ if $\frac{1-F(\beta^*)}{1-F(\bar{\beta})} \geq \rho_2^*$.

The initial beliefs of children of giving up parents are the same as they were without empathy, thus they cannot do better than play according to S_2^G. Recall that if condition (7.13) holds, these children will attempt willpower, persevere if $\beta_{2i} \geq \beta''$, and select willpower with probability p_2 in the second period upon observing P. Otherwise, they do not attempt willpower.

Step 3. Find the expected children's utility when the strategy profile $\{S_1^e; S_2^P, S_2^G\}$ is played.

Mixed Equilibrium. Consider first the case in which children of giving up parents have sufficient self-confidence to attempt willpower.

If a strong type parent perseveres, the parent's expectation of his offspring's utility, provided child plays according to S_2^P, will be:

$$\begin{aligned} E[u_{2i}(S^{2P}) \mid \beta_{1i} \geq \beta^*, P] &= (\theta + (1-\theta)(1-F(\beta^*)))(B-c+\delta(B-c)) \\ &+(1-\theta)(F(\beta^*)-F(\hat{\beta}))(B-c+\delta b) + (1-\theta)F(\hat{\beta})(b+\delta a) \end{aligned} \tag{8.4}$$

If a strong type parent gives up, the parent's expectation of his offspring's utility, according to S_2^G, will be:

$$E[u_{2i}(S^{2G}) \mid \beta_{1i} \geq \beta^*, G] = (\theta + (1-\theta)(1-F(\beta^*)))(B - c + \delta(p_2(B-c) + (1-p_2)a)) \tag{8.5}$$
$$+(1-\theta)(F(\beta^*) - F(\beta''))(B - c + \delta(p_2 b + (1-p_2)a)) + (1-\theta)F(\beta'')(b + \delta a)$$

From (8.4) and (8.5), it is clear that $E[u_{2i}(S^{2G}) \mid \beta_{1i} \geq \beta^*, G] < E[u_{2i}(S^{2P}) \mid \beta_{1i} \geq \beta^*, P]$. The first term on the RHS of (8.5) is smaller than the first term on RHS of (8.4) as $B - c + \delta(p_2 b + (1-p_2)a) < B - c + \delta(B-c)$. The sum of the second and the third terms on the RHS of (8.5) is also smaller than the corresponding sum in (8.4) as the former attaches a higher weight to a smaller payoff, $F(\beta'') > F(\hat{\beta})$ and $b + \delta a < B - c + \delta(p_2 b + (1-p_2)a) < B - c + \delta b$. Thus, it remains the dominant strategy for the strong type to persevere at $t = 1$, $\tau = 2$.

If an intermediate parent with $\beta_{1i} \in [\hat{\beta}, \beta^*)$ perseveres, the expected utility of his child, according to S_2^P, will be:

$$E(u_{2i}(S^{2P}) \mid \beta_{1i} \in [\hat{\beta}, \beta^*), P) = (1-\theta)(1-F(\beta^*))(B - c + \delta(B-c)) \tag{8.6}$$
$$+(\theta + (1-\theta)(F(\beta^*) - F(\hat{\beta})))(B - c + \delta b) + (1-\theta)F(\hat{\beta})(b + \delta a)$$

And if such parent gives up:

$$E(u_{2i}(S^{2G}) \mid \beta_{1i} \in [\hat{\beta}, \beta^*), G) = (1-\theta)(1-F(\beta^*))(B - c + \delta(p_2(B-c) + (1-p_2)a)) \tag{8.7}$$
$$+(\theta + (1-\theta)(F(\beta^*) - F(\beta'')))(B - c + \delta(p_2 b + (1-p_2)a)) + (1-\theta)F(\beta'')(b + \delta a)$$

By the same logic as with the children of strong types, it is clear from (8.6) and (8.7) that $E(u_{2i}(S^{2P}) \mid \beta_{1i} \in [\hat{\beta}, \beta^*), P) > E(u_{2i}(S^{2G}) \mid \beta_{1i} \in [\hat{\beta}, \beta^*), G)$. Therefore, it remains optimal for parents with $\beta_{1i} \geq \hat{\beta}$ to persevere at $t = 1, \tau = 2$ and moreover, the last parent to persevere must be below the $\hat{\beta}$ threshold, i.e. $\bar{\beta} < \hat{\beta}$.

Then, for parents with $\beta_{1i} \in [\bar{\beta}, \hat{\beta})$, who would not have persevered without empathy, the expectation of children's utility will be:

$$E(u_{2i}(S^{2P}) \mid \beta_{1i} \in [\bar{\beta}, \hat{\beta}), P) = (1-\theta)(1-F(\beta^*))(B - c + \delta(B-c)) \tag{8.8}$$
$$+(1-\theta)(F(\beta^*) - F(\hat{\beta}))(B - c + \delta b) + (\theta + (1-\theta)F(\hat{\beta}))(b + \delta a)$$

And if $\beta_{1i} \in [\bar{\beta}, \hat{\beta})$ parents give up:

$$E(u_{2i}(S^{2G}, G) \mid \beta_{1i} \in [\bar{\beta}, \hat{\beta})) = (1-\theta)(1-F(\beta^*))(B-c+\delta(p_2(B-c)+(1-p_2)a)) \tag{8.9}$$
$$+(1-\theta)(F(\beta^*)-F(\beta''))(B-c+\delta(p_2 b+(1-p_2)a)) + (\theta+(1-\theta)F(\beta''))(b+\delta a)$$

Once again, $E(u_{2i}(S^{2P}) \mid \beta_{1i} \in [\bar{\beta}, \hat{\beta}), P) > E(u_{2i}(S^{2G}) \mid \beta_{1i} \in [\bar{\beta}, \hat{\beta}), G)$. It is also the case that $E(u_{2i}(S^{2P}) \mid \beta_{1i} < \bar{\beta}, P) = E(u_{2i}(S^{2P}) \mid \beta_{1i} \in [\bar{\beta}, \hat{\beta}), P)$, even though this is off the equilibrium path, and $E(u_{2i}(S^{2G}) \mid \beta_{1i} < \bar{\beta}, G) = E(u_{2i}(S^{2G}) \mid \beta_{1i} \in [\bar{\beta}, \hat{\beta}), G)$.

To determine $\bar{\beta}$, we compute the expected gain in child welfare from the $\bar{\beta}$ parent persevering:

$$\begin{aligned} & E(u_{2i}(S^{2P}) \mid \bar{\beta}, P) - E(u_{2i}(S^{2G}) \mid \bar{\beta}, G) = \\ &= (1-\theta)\delta(1-p_2)[(1-F(\beta^*))(B-c-a) + (F(\beta^*)-F(\beta''))(b-a)] \\ & \quad +(1-\theta)(F(\beta'')-F(\hat{\beta}))(B-c-b+\delta(b-a)) \end{aligned} \tag{8.10}$$

Therefore,

$$\bar{\beta} = c\{B-b+\delta(b-a)+\mu(1-\theta)[\delta(1-p_2)[(1-F(\beta^*))(B-c-a)+(F(\beta^*)-F(\beta''))(b-a)]+ \tag{8.11}$$
$$(F(\beta'')-F(\hat{\beta}))(B-c-b+\delta(b-a))]\}^{-1}$$

This $\bar{\beta}$ identifies the lowest parent who is willing to persevere if willpower is played with probability one following perseverance in period two. All agents with $\beta < \bar{\beta}$ will give up at $t=1$, $\tau=2$.

Then at $t=1$, $\tau=1$, parents will attempt willpower if the expected payoff to W exceeds the expected payoff to NW :

$$\begin{aligned} & (1-F(\beta^*))(B-c+\delta(B-c)+\mu E(u_{2i}(S^{2P}) \mid \beta_{1i} \geq \beta^*, P) \\ & +(F(\beta^*)-F(\hat{\beta}))(B-c+\delta b+\mu E(u_{2i}(S^{2P}) \mid \beta_1 \in [\hat{\beta}, \beta^*), P) \\ & +(F(\hat{\beta})-F(\bar{\beta}))(B-c+\delta b+\mu E(u_{2i}(S^{2P}) \mid \beta_1 \in [\bar{\beta}, \hat{\beta}), P) \\ & +F(\bar{\beta})(b+\delta a+\mu E(u_{2i}(S^{2G}) \mid \beta_{1i} < \bar{\beta}, G) \geq \tfrac{a}{\gamma}+\delta a+\mu E(u_c^{NW}) \end{aligned} \tag{8.12}$$

where $E(u_c^{NW})$ is the expected utility of children, whose parents did not attempt willpower.

Condition (8.12) can also be written as:

$$(1-F(\beta^*))(B-c+\delta(B-c))+(F(\beta^*)-F(\bar{\beta}))(B-c+\delta b)+F(\bar{\beta})(b+\delta a) \quad (8.13)$$
$$+\mu(1-\theta)\left[\begin{array}{c} (1-F(\bar{\beta}))\left[\begin{array}{c}(1-F(\beta^*))(B-c+\delta(B-c)) \\ +(F(\beta^*)-F(\hat{\beta}))(B-c+\delta b)+F(\hat{\beta})(b+\delta a)\end{array}\right] \\ +F(\bar{\beta})\left[\begin{array}{c}(1-F(\beta^*))(B-c+\delta(p_2(B-c)+(1-p_2)a)) \\ +(F(\beta^*)-F(\beta''))(B-c+\delta(p_2 b+(1-p_2)a))+F(\beta'')(b+\delta a)\end{array}\right]\end{array}\right]$$
$$+\mu\theta\left[(1-F(\beta^*))(B-c+\delta(B-c))+(F(\beta^*)-F(\hat{\beta}))(B-c+\delta b)+F(\hat{\beta})(b+\delta a)\right]$$
$$\geq \tfrac{a}{\gamma}+\delta a+\mu E(u_c^{NW})$$

It remains to determine $E(u_c^{NW})$. If the information node in which no parents attempt willpower at $t=1$, $\tau=1$ is reached, the equilibrium is a pooling one. If all parents choose NW, no additional information about parent's type is revealed to the children. Therefore, children would have the same initial beliefs as their parents; then it must be the case that if parents found it optimal to choose NW, so will the children if they have the same initial beliefs. Thus, $E(u_c^{NW})=(1+\delta)a$.

Pure Equilibrium. Finally, consider the case in which the self-confidence of children of giving up parents is not sufficient for them to attempt willpower. In this case, the expected utility of a child of a giving up parent, regardless of parent's type, would be:

$$E(\underline{\mathrm{u}}_c^G)=(1+\delta)a \quad (8.14)$$

According to Assumption 1, some self-control is better than none, thus in terms of payoffs, this is the worst outcome for the children. This decline in children's utility from parents' giving up should provide an extra incentives for the parents to persevere, and so the last intermediate parent to persevere in this case will be some $\bar{\beta}'<\bar{\beta}$. ■

8.2 *Proof of Proposition 3.7.*

Proof. Suppose agents in all generations play a stationary strategy profile $\{S^{eP}, S^{eG}\}$. If we can find such β'', $\tilde{\beta}$ and p_2^e for which $\{S^{eP}, S^{eG}\}$ is sequentially rational and show that beliefs are consistent with $\{S^{eP}, S^{eG}\}$ in each generation, then we will have a Perfect Bayesian Equilibrium.

Step 1. Take any generation k. Behaviour in period two is the same as in the two-generation case: only the strong types persevere at the end of the game and only those agents with the updated second period beliefs above the ρ_2^* threshold attempt willpower at $t=2$, $\tau=1$, at least with some probability. Consider what happens at $t=1$, $\tau=2$ when $\{S^{eP}, S^{eG}\}$ is played. For children of persevering parents with $\beta_{ki}\geq\beta^*$, persevering yields $B-c/\beta_{ki}+\delta(B-c)+\mu V_{k+1,i}(\beta_{ki}\geq\beta^*,P)$, and giving up yields $b + \delta a + \mu V_{k+1,i}(\beta_{ki} \geq \beta^*, G)$. Similarly, for children of persevering parents with $\beta_{ki} < \beta^*$, persevering yields $B - c/\beta_{ki} + \delta b + \mu V_{k+1,i}(\beta_{ki} < \beta^*, P)$, and giving up yields $b + \delta a + \mu V_{k+1,i}(\beta_{ki} < \beta^*, G)$. By continuity, there exists an agent $\tilde{\beta}_{ki}$ who is indifferent between persevering and giving up:

$$\tilde{\beta}_{ki} = \frac{c}{B-b+\delta(b-a)+\mu[V_{k+1,i}(\tilde{\beta},P)-V_{k+1,i}(\tilde{\beta},G)]} \tag{8.15}$$

To find $\tilde{\beta}_{ki}$, we need to find $V_{k+1,i}(\tilde{\beta}, P) - V_{k+1,i}(\tilde{\beta}, G)$. If $V_{k+1,i}(\tilde{\beta}, P) - V_{k+1,i}(\tilde{\beta}, G) > 0$, then $\tilde{\beta}_{ki} < \hat{\beta}$, and all children of persevering parents with $\beta_{kj} \geq \tilde{\beta}_{ki}$ will have the incentive to persevere at $t = 2$, $\tau = 1$, as long as their updated second period beliefs satisfy the informativeness constraint, i.e. $\frac{\theta\frac{1-F(\beta^*)}{\lambda_{k-1}}+(1-\theta)(1-F(\beta^*))}{1-F(\tilde{\beta})} \geq \rho_2^*$, where λ_{k-1} is the proportion of agents persevering in generation $k-1$.

For children of giving up parents, if $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\tilde{\beta})} \geq \rho_2^*$, they play the same strategy as children of persevering parents. If, $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\tilde{\beta})} < \rho_2^*$, then the last agent to persevere at $t = 1$, $\tau = 2$ is β'', which solves $\frac{(1-\theta)(1-F(\beta^*))}{1-F(\beta'')} = \rho_2^*$. Since the threshold ρ_2^* is unchanged, β'' is the same in this equilibrium as in the two generations and no empathy case. For S^G to be deviation-proof, β'' must be the last child of giving-up parents who has an incentive to persevere, i.e. he must be indifferent between persevering, which yields him $B - c/\beta'' + \delta(p_2^e b + (1-p_2^e)a) + \mu V_{k+1}(\beta'', P)$, and giving up, which yields $b + \delta a + \mu V_{k+1}(\beta'', G)$. Therefore, p_2^e must be:

$$p_2^e = \frac{b-B+c/\beta''-\mu[V_{k+1}(\beta'',P)-V_{k+1}(\beta'',G)]}{\delta(b-a)} \tag{8.16}$$

To find p_2^e, we must find $V_{k+1}(\beta'', P) - V_{k+1}(\beta'', G)$. Notice that if $V_{k+1}(\beta'', P) - V_{k+1}(\beta'', G) > 0$, $p_2^e < p_2$, i.e. the second period probability of choosing willpower for children of giving up parents is actually lower under empathy.

Step 2. We turn now to determining the expected utility of offsprings, $V_{k+1,i}(\beta_{ki}, A_{ki})$. Since agents in all generations are assumed to play their best response, we refer to the expected utility as the expected value functions. Firstly, if a stationary equilibrium strategy is played in the infinite generation context, then $V_{k+1}(\beta_k, A^k_{t=1,\tau=2}) = V_k(\beta_{k-1}, A^{k-1}_{t=1,\tau=2}) \equiv V(\beta, A_{t=1,\tau=2})$ $\forall i, k$, as long as $\beta_k = \beta_{k-1} = \beta$ and $A^k_{t=1,\tau=2} = A^{k-1}_{t=1,\tau=2} = A_{t=1,\tau=2}$, i.e. the expected value function for future generations is independent of a particular generation number; it is a function of a state, defined by two variables - the type of parent, β, and the $t = 1$, $\tau = 2$ action of the parent, A.

Then we can write the expected utility, or value function, of a child of $\tilde{\beta}$ parent who had persevered as:

$$V(\tilde{\beta}, P) = (1-\theta)(1-F(\beta^*))(B - c + \delta(B-c) + \mu V(\beta \geq \beta^*, P)) \\ +(\theta + (1-\theta)(F(\beta^*) - F(\tilde{\beta})))(B - c + \delta b + \mu V(\tilde{\beta} \leq \beta < \beta^*, P)) \\ +(1-\theta)F(\tilde{\beta})(b + \delta a + \mu V(\beta < \tilde{\beta}, G)) \tag{8.17}$$

where $V(\beta \geq \beta^*, P))$ is the value of being a child of a strong and persevering parent, $V(\tilde{\beta} \leq \beta < \beta^*, P))$ is the value of a child of an intermediate parent with $\tilde{\beta} \leq \beta < \beta^*$, who had also persevered, and $V(\beta < \tilde{\beta}, G)$ is the value function for the child of a weak giving up parent. Then the value function for a child of a $\tilde{\beta}$ parent who had given up is:

$$V(\tilde{\beta}, G) = (1-\theta)(1-F(\beta^*))(B - c + \delta(p_2^e(B-c) + (1-p_2^e)a) + \mu V(\beta \geq \beta^*, P)) \tag{8.18}$$
$$+(1-\theta)(F(\beta^*) - F(\beta''))(B - c + \delta(p_2^e b + (1-p_2^e)a) + \mu V(\beta'' \leq \beta < \beta^*, P)) \\ +(\theta + (1-\theta)F(\beta''))(b + \delta a + \mu V(\beta < \beta'', G))$$

Similarly, for a child of β'' parent who had persevered, the value function can be written as:

$$V(\beta'', P) = (1-\theta)(1-F(\beta^*))(B - c + \delta(B-c) + \mu V(\beta \geq \beta^*, P)) \\ +(\theta + (1-\theta)(F(\beta^*) - F(\tilde{\beta})))(B - c + \delta b + \mu V(\tilde{\beta} \leq \beta < \beta^*, P)) \\ +(1-\theta)F(\tilde{\beta})(b + \delta a + \mu V(\beta < \tilde{\beta}, G)) \tag{8.19}$$

And if the β'' had given up:

$$V(\beta'', G) = (1-\theta)(1-F(\beta^*))(B - c + \delta(p_2^e(B-c) + (1-p_2^e)a) + \mu V(\beta \geq \beta^*, P)) \tag{8.20}$$
$$+(\theta + (1-\theta)(F(\beta^*) - F(\beta'')))(B - c + \delta(p_2^e b + (1-p_2^e)a) + \mu V(\beta'' \leq \beta < \beta^*, P)) \\ +(1-\theta)F(\beta'')(b + \delta a + \mu V(\beta < \beta'', G))$$

Conditional on parent's action, the value function is a step function of the parent's type. It is easy to see that

$$V(\beta \geq \beta^*, P) > V(\tilde{\beta} \leq \beta < \beta^*, P) = V(\beta'' \leq \beta < \beta^*, P) \tag{8.21}$$

and

$$V(\beta \geq \beta^*, G) > V(\beta'' \leq \beta < \beta^*, G) > V(\tilde{\beta} \leq \beta < \beta^*, G) = V(\beta < \beta'', G) = V(\beta < \tilde{\beta}, G) \tag{8.22}$$

With this in mind, we can write:

$$V(\beta'', P) - V(\beta'', G) = \delta(1 - p_2^e)\left[(1-\theta)(1 - F(\beta^*))(B - c - a) + (1-\theta)(F(\beta^*) - F(\beta'')(b - a)\right] \tag{8.23}$$

$$+(1-\theta)(F(\beta'') - F(\tilde{\beta}))(B - c - b + \delta(b-a) + \mu(V(\beta'', P) - V(\beta'', G)) + \theta\delta(1 - p_2^e)(b - a)$$

$$V(\tilde{\beta}, P) - V(\tilde{\beta}, G) = \delta(1 - p_2^e)\left[(1-\theta)(1 - F(\beta^*))(B - c - a) + (1-\theta)(F(\beta^*) - F(\beta'')(b - a)\right] \tag{8.24}$$

$$+(1-\theta)(F(\beta'') - F(\tilde{\beta}))(B - c - b + \delta(b-a) + \mu(V(\tilde{\beta}, P) - V(\tilde{\beta}, G))$$
$$+\theta(B - c - b + \delta(b-a) + \mu(V(\tilde{\beta}, P) - V(\tilde{\beta}, G))$$

$$p_2^e = \frac{b - B + c/\beta'' - \mu[V(\beta'', P) - V(\beta'', G)]}{\delta(b-a)} \tag{8.25}$$

Solving equations (8.24) and (8.25) simultaneously to find p_2^e, and substituting back into equation (8.23) gives:

$$V(\beta'', P) - V(\beta'', G) = (B - b - c/\beta'' + \delta(b-a))\frac{D}{1-\mu D} + (1-\theta)(F(\beta'') - F(\tilde{\beta}))c\frac{1-\beta''}{\beta''(1-\mu D)} \tag{8.26}$$

$$V(\tilde{\beta}, P) - V(\tilde{\beta}, G) = (B - b - c/\beta'' + \delta(b-a))\frac{D}{1-\mu D} \tag{8.27}$$
$$+(1-\theta)(F(\beta'') - F(\tilde{\beta}))c\frac{1-\beta''}{\beta''(1-\mu D)} + \frac{c\theta(1-\beta'')}{\beta''(1-\mu\theta-\mu(1-\theta)(F(\beta'')-F(\tilde{\beta})))}$$

$$1 - p_2^e = \frac{B - b - c/\beta'' + \delta(b-a) + \mu(1-\theta)(F(\beta'') - F(\tilde{\beta}))c\frac{1-\beta''}{\beta''}}{\delta(b-a)(1-\mu D)} \tag{8.28}$$

where $D \equiv \theta + (1-\theta)(F(\beta^*) - F(\tilde{\beta})) + (1-\theta)(1 - F(\beta^*))\frac{B-c-a}{b-a}$.
Then use equation (8.27) to find $\tilde{\beta}$:

$$\tilde{\beta} = c\left[\frac{B - b - c/\beta'' + \delta(b-a)}{1-\mu D} + \frac{c}{\beta''}\left(1 + (1-\theta)(F(\beta'') - F(\tilde{\beta}))\frac{1-\beta''}{(1-\mu D)}\right) + \frac{\theta(1-\beta'')}{(1-\mu\theta-\mu(1-\theta)(F(\beta'')-F(\tilde{\beta})))}\right]^{-1} \tag{8.29}$$

By the same argument as in the case of two generations, $V(\beta, P) - V(\beta, G) \geq 0$ for all β ($V(\beta, P)$ attaches higher weight to higher payoffs), therefore, $\tilde{\beta} < \hat{\beta}$ and $p_2^e < p_2$.

Step3. We now find the proportion of the population that perseveres in the long run. Since all generations play a stationary strategy, this is equivalent to the case of infinite generations without empathy, but with $\tilde{\beta}$ instead of $\hat{\beta}$. Therefore,

$$\lambda_k^e = (1 - F(\tilde{\beta}))[\theta + (1-\theta)(F(\beta'') - F(\hat{\beta}))]^{k-1} + \tfrac{1-F(\beta'')}{1-(F(\beta'')-F(\hat{\beta}))} \tag{8.30}$$

and

$$\lim_{k\to\infty} \lambda_k^e = \tfrac{1-F(\beta'')}{1-(F(\beta'')-F(\tilde{\beta}))}$$

and $\lim_{k\to\infty} \lambda_k^e > \lim_{k\to\infty} \lambda_k$.

It is clear that $\frac{1-F(\beta'')}{1-(F(\beta'')-F(\tilde{\beta}))} < 1 - F(\tilde{\beta})$, so that the informativeness constraint is always satisfied for children of persevering parents.

Step 4. Condition (5.4) guarantees that children of giving-up parents have sufficient self-confidence to attempt willpower in the first period. It is possible to calculate the value functions explicitly, although for the sake of saving space, we leave the condition (5.4) in its value function form, which also simplifies the exposition. If condition (5.4) did not hold, so that children attempted no willpower at the beginning of the game, there could still be a pure equilibrium of the kind described in multiple generations without empathy case.

It is also clear that under the assumption $\frac{1-F(\beta^*)}{1-F(\tilde{\beta})} \geq \rho_2^*$, generation one have sufficiently high initial and posterior beliefs to play the same strategy as the children of persevering parents. ■

8.3 *Proof of Proposition 3.8*

Proof. Suppose the strategy profile, $\{S_1^{LP}; S_2^{NW}, S_2^G, S_2^P\}$ is played. We need to show that this strategy profile is sequentially rational given beliefs, and that beliefs are correct in equilibrium.

Step 1: Consider the parents' problem at $t = 2$, $\tau = 2$. For a β_{1i} parent, persevering yields $B - c/\beta_{1i} + \mu E[u_{2i}^{LP}(S^{2P}) \mid \beta_{1i}, P]$, and giving up yields $b + \mu E[u_{2i}^{LP}(S^{2G}) \mid \beta_{1i}, G]$. By continuity, there exists some β^{**} parent who is indifferent between P and G at $t = 2$, $\tau = 2$:

$$\beta^{**} = \frac{c}{B - b + \mu(E[u_{2i}^{LP}(S^{2P}) \mid \beta^{**}, P] - E[u_{2i}^{LP}(S^{2G}) \mid \beta^{**}, G])} \tag{8.31}$$

All parents with $\beta_{1i} \geq \beta^{**}$ will find it optimal to persevere at $t = 2$, $\tau = 2$. Then at $t = 2$, $\tau = 1$, choosing willpower yields $\rho_2(B - c + \mu E[u_{2i}^{LP}(S^{2P}) \mid \beta_{1i} \geq \beta^{**}, P]) + (1-\rho_2)(b + \mu E[u_{2i}^{LP}(S^{2G}) \mid \beta_{1i} < \beta^{**}, G])$; and not attempting willpower yields $\alpha/\gamma + \mu E[u_{2i}^{LP}(S^{NW}) \mid \rho_2, NW]$, where $\rho_2 = \Pr(\beta_{1i} \geq \beta^{**} \mid A_{t=1,\tau=2}^{1i}, \rho_{t=1})$. Then, parents will choose to attempt willpower at $t = 2$, $\tau = 1$ if their updated second period beliefs satisfy the new informativeness constraint: $\Pr(\beta_{1i} \geq \beta^{**} \mid A^{1i}_{t=1,\tau=2}, \rho_{t=1}) \geq \rho_2^{**}$, where

$$\rho_2^{**} = \frac{a/\gamma - b + \mu(E[u_{2i}^{LP}(S^{NW}) | \rho_2^{**}, NW] - E[u_{2i}^{LP}(S^{2G}) | \beta_{1i} < \beta^{**}, G])}{B - c - b + \mu(E[u_{2i}^{LP}(S^{2P}) | \beta_{1i} \geq \beta^{**}, P] - E[u_{2i}^{LP}(S^{2G}) | \beta_{1i} < \beta^{**}, G])} \tag{8.32}$$

The threshold level of self-confidence, ρ^{**}, makes the parent indifferent between W and NW at $t = 2$, $\tau = 1$, taking into account the expected welfare of the children.

Then, at $t = 1$, $\tau = 2$, under the strategy profile $\{S_1^{LP}; S_2^{NW}, S_2^G, S_2^P\}$, giving up yields $b + \delta(a + \mu E[u_{2i}^{LP}(S^{NW}) \mid \beta_{1i.}, NW)$ for any $\beta_{1i.}$ On the other hand, persevering yields $B - c/\beta_{1i} + \delta(B - c + \mu E[u_{2i}^{LP}(S^{2P}) \mid \beta_{1i} \geq \beta^{**}, P])$ for $\beta_{1i} \geq \beta^{**}$ and $B - c/\beta_{1i} + \delta(b + \mu E[u_{2i}^{LP}(S^{2G}) \mid \beta_{1i} < \beta^{**}, G])$ for $\beta_{1i} < \beta^{**}$. Suppose $\beta^{**} > \hat{\beta}$, then the last parent who has an incentive to persevere is

$$\bar{\beta}^{LP} = \frac{c}{B - b + \delta(b - a) + \delta\mu(E[u_{2i}^{LP}(S^{2G}) \mid \bar{\beta}^{LP}, G] - E[u_{2i}^{LP}(S^{NW}) \mid \bar{\beta}^{LP}, NW])} \tag{8.33}$$

For the purpose of generating a pure strategy equilibrium in generation 1, we assume that $\frac{1-F(\beta^*)}{1-F(\bar{\beta}^{LP})} \geq \rho_2^{**}$. Then, all parents with $\beta_{1i} \geq \bar{\beta}^{LP}$ persevere at $t = 1$, $\tau = 2$, parents with $\beta_{1i} < \bar{\beta}^{LP}$give up, and no-one has an incentive to deviate.

From (8.33), $\bar{\beta}^{LP} < \hat{\beta}$ if $E[u_{2i}^{LP}(S^{2G}) \mid \bar{\beta}^{LP}, G] > E[u_{2i}^{LP}(S^{NW}) \mid \bar{\beta}^{LP}, NW]$. Below we find the relative sizes of children's expected utilities given parental actions, and therefore find the relative magnitude$\bar{\beta}^{LP}$ and β^{**} compared to $\hat{\beta}$, and of ρ_2^{**} compared to ρ_2^*.

Step 2: Consider the initial beliefs of generation two. Since generation two do not themselves have children, they play according to the same strategy as the no-altruism case. Once again, children of persevering parents have higher initial beliefs than their parents, which implies they have sufficient self-confidence to attempt willpower and persevere, conditional on type, whenever their parents did so:

$$\rho_{LP,1}^{2|P} = \Pr(\beta_{2i} \geq \beta^* \mid A^{1i}_{t=2} = P) = \theta\tfrac{1-F(\beta^*)}{1-F(\beta^{**})} + (1-\theta)(1 - F(\beta^*)) \tag{8.34}$$

Therefore, children of persevering parents cannot do better than play according to S_2^P, and their expected utility is the same as in Proposition 3.6, $E[u_{2i}^{LP}(S^{2P}) \mid \beta_{1i}, P] = E[u_{2i}(S^{2P}) \mid \beta_{1i}, P] \equiv E(u_{2i}^P \mid \beta_{1i})$.

The initial beliefs of children of giving up parents are:

$$\rho_{LP,1}^{2|G} = \Pr(\beta_{2i} \geq \beta^* \mid A^{1i}_{t=2} = G) = (1-\theta)(1 - F(\beta^*)) \tag{8.35}$$

Since children can deduce that, according to S_1^{LP}, only parents with $\beta_{1i} < \bar{\beta}^{LP}$, who had given up in the first period, do not attempt willpower in the second period, the initial beliefs of children whose parents did not attempt willpower are the same as the initial beliefs of children whose parents gave up at the end of the second period:

$$\rho_{LP,1}^{2|NW} = \Pr(\beta_{2i} \geq \beta^* \mid A_{t=2}^{1i} = NW) = (1-\theta)(1-F(\beta^*)) = \rho_{LP,1}^{2|G} \tag{8.36}$$

Since children of giving up and of not attempting willpower parents have the same initial beliefs, (8.35) and(8.36), they cannot do better than play according to the strategy S_2^G, and therefore the expected utility of such children is again as in Proposition 3.6, $E[u_{2i}^{LP}(S^{2G}) \mid \beta_{1i}, G] = E[u_{2i}^{LP}(S^{NW}) \mid \beta_{1i.}, NW) = E(u_{2i}^G \mid \beta_{1i})$.

However, we have already shown in the proof of Proposition (3.6) that if children play according to $\{S_2^P, S_2^G\}$, then $E(u_{2i}^P \mid \beta_{1i}) > E(u_{2i}^G \mid \beta_{1i})$ for all β_{1i}.

Therefore, from (8.33), we have:

$$\bar{\beta}^{LP} = \frac{c}{B-b+\delta(b-a)} = \hat{\beta} \tag{8.37}$$

and from (8.32):

$$\rho_2^{**} = \frac{a/\gamma - b}{B-c-b+\mu(E[u_{2i}^{LP}(S^{2P})|\beta_{1i} \geq \beta^{**}, P] - E[u_{2i}^{LP}(S^{2G})|\beta_{1i} < \beta^{**}, G])} < \rho_2^* \tag{8.38}$$

And,

$$\beta^{**} = \frac{c}{B-b+\mu(E[u_{2i}^{LP}(S^{2P}) \mid \beta^{**}, P] - E[u_{2i}^{LP}(S^{2G}) \mid \beta^{**}, G])} < \beta^* \tag{8.39}$$

Finally, at $t=1$, $\tau=1$, parents attempt willpower if:

$$(1-F(\beta^{**}))(B-c+\delta(B-c+\mu E[u_{2i}^{LP}(S^{2P}) \mid \beta_{1i} \geq \beta^{**}, P])) \tag{8.40}$$

$$+(F(\beta^{**})-F(\hat{\beta}))(B-c+\delta(b+\mu E[u_{2i}^{LP}(S^{2G}) \mid \hat{\beta} \leq \beta_{1i} < \beta^{**}, G])) \tag{8.41}$$

$$+F(\hat{\beta})(b+\delta a+\mu E[u_{2i}^{LP}(S^{NW}) \mid \beta_{1i.} < \hat{\beta}, NW)) \geq \tfrac{a}{\gamma} + \delta a + \mu E(u_c^{NW1})$$

As before, we can assume that $E(u_c^{NW1}) = (1+\delta)a$, and so for any 'reasonable' payoffs, condition (8.40) should hold trivially.

Recall that we assumed that $\beta^{**} > \hat{\beta}$, however it is possible that for $\mu(E[u_{2i}^{LP}(S^{2P}) \mid \beta^{**}, P] - E[u_{2i}^{LP}(S^{2G}) \mid \beta^{**}, G]) > \delta(b-a)$, $\beta^{**} < \hat{\beta}$. This happens if:

$$\mu[(1-\theta)\delta(1-p_2)[(1-F(\beta^*))(B-c-a)+(F(\beta^*)-F(\beta''))(b-a)]$$
$$+(1-\theta)(F(\beta'')-F(\hat{\beta}))(B-c-b+\delta(b-a))] > \delta(b-a)$$

In this case, all parents with $\beta_{1i} \geq \beta^{**}$ persevere in both periods, as long as $\frac{1-F(\beta^*)}{1-F(\beta^{**})} \geq \rho_2^*$. Equilibrium strategies remain unchanged otherwise. ■

References

[1] Ainslie, George, (2001), *Breakdown of Will*, Cambridge University Press

[2] Ameriks, John, Caplin, Andrew; and Leahy, John J, (2003), "Wealth Accumulation and the Propensity to Plan", *Quarterly Journal of Economics*, 118, pp. 1007—1048

[3] Battaglini, Marco, Roland Benabou and Jean Tirole, (2005), "Self-Control in Peer Groups," *Journal of Economic Theory*, 123, pp.105-134

[4] Baumeister, Roy F., Engels, Rutger C. M. E., and Finkenaur, Catrin, (2005), "Parenting Behaviour and Adolescent Behavioural and Emotional Problems: The Role of Self Control," *International Journal of Behavioural Development*, vol 29 (1), pp. 58-69

[5] Becker, Gary S. and Kevin M. Murphy, (1988), "A Theory of Rational Addiction," *Journal of Political Economy*, 96, (Aug., 1988), pp.675-700

[6] Becker, Gary S. and Tomes, N. (1979), "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 87(6)„ pp. 1153-1189

[7] Benhabib, Jess and Alberto Bisin, (2005), "Modelling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions," *Games and Economic Behaviour*, vol 52, pp. 460-492

[8] Benabou, Roland and Jean Tirole, (2002), "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117, (Aug., 2002), pp. 871-915

[9] Benabou, Roland and Jean Tirole, (2003), "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70, (2003), pp.489-520

[10] Benabou, Roland and Jean Tirole, (2004), "Willpower and Personal Rules," *Journal of Political Economy*, University of Chicago Press, vol. 112(4), pp. 848-886

[11] Benabou, Roland and Jean Tirole, (2006), "Identity, Dignity and Taboos: Beliefs as Assets," *IZA Discussion Papers* 2583, (2006)

[12] Benhabib, Jess and Bisin, Alberto, (2006), "The Distribution of Wealth and Redistributive Policies," UCLA Department of Economics, Working Paper Archive, number 122247000000001162

[13] Bernheim, B. Douglas, (1994), "A Theory of Conformity," *Journal of Political Economy*, 102, pp.841-877

[14] Bisin, Alberto and Thierry Verdier, (2001), "Agents with imperfect empathy may survive natural selection," *Economics Letters*, 71, pp. 277-285

[15] Bisin, Alberto and Thierry Verdier, (2001), "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 97, (2001), pp.298-319

[16] Black, Sandra E. & Devereux, Paul & Salvanes, Kjell G., (2008), "Like Father, Like Son? A Note on the Intergenerational Transmission of IQ Scores," IZA Discussion Papers 3651, Institute for the Study of Labor (IZA)

[17] Blanden, Jo, Goodman, Alissa, Gregg, Paul and Machin, Stephen, (2002), "Changes in Intergenerational Mobility in Britain," *CEP Discussion Paper* number 0517

[18] Bodner, Ronit and Drazen Prelec, (1997), "The Diagnostic Value of Actions in a Self-Signalling Model," *Manuscript*, Massachuesetts Institute of Technology

[19] Bodner, Ronit and Drazen Prelec, (1997), "Self-Signaling and Diagnostic Utility in Everyday Decision Making," *Unpublished Manuscript*, Massachuesetts Institute of Technology

[20] Bodner, Ronit and Drazen Prelec, (2003), "Self Signalling and Self Control," in *Time and Decision*, Loewenstein G. and Read D., and Baumeister R. F. (eds), Russell Sage Press, New York

[21] Bowles, Samuel and Herbert Gintis (2001), "The Inheritance of Economic Status: Education, Class and Genetics," in Marcus Feldman and Paul Baltes (eds.) *International Encyclopedia of the Social and Behavioral Sciences: Genetics, Behavior and Society*, New York: Oxford University Press and Elsevier

[22] Bouchard, T. J., Jr. and M. McGue (1981), "Familial Studies of Intelligence," *Science* ,212, pp. 1055–1059

[23] Brocas, Isabelle and Carrillo, Juan D., (2003), "Biases in Perceptions, Beliefs and Behaviours," *USC CLEO Research Paper* No. C04-19

[24] Brocas, Isabelle and Carrillo, Juan D., (2005), "Biased Decision Making by Agents with Unbiased Beliefs," *Manuscript*

[25] Camerer, Colin F., Linda Babcock, George Loewenstein and Richard H. Thaler, (1997), "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quarterly Journal of Economics*, 112, pp. 407-41

[26] Caplin, A. and Leahy, J. (2004), "The Social Discount Rate," *Journal of Political Economy*, 112, pp. 1257-1268

[27] Carrillo, Juan D., (2005), "To be consumed with moderation," *European Economic Review*, 49, pp.99-111

[28] Carrillo, Juan D. and Thomas Mariotti, (2000), "Strategic Ignorance as a Self-Disciplining Device," *Review of Economic Studies*, 67, pp. 529-544

[29] Charles, Kerwin Kofi and Hurst, Eric, (2003), "The Correlation of Wealth Across Generations," *Journal of Political Economy*, vol 11(6)

[30] Fishbach, Ayelet and Trope, Yaacov, (2000), "Counteractive Self-Control in Overcoming Temptation," *Journal of Personality and Social Psychology*, vol 79(4), pp. 493-506

[31] Glazer, Amihai and Konrad, Kai A., (1996), "A Signaling Explanation for Charity," *The American Economic Review*, vol 86 (4), pp. 1019-1028

Gul, Faruk and Wolfgang Pesendorfer, (2001), "Temptation and Self-Control," *Econometrica*, vol. 69, no. 6, pp. 1403-1435

[32] Gul, Faruk and Wolfgang Pesendorfer (2004), "Self Control, Revealed Preference and Consumption Choice," *Review of Economic Dynamics*, vol. 7(2), pp. 243-264

[33] Frederick, Shane, Loewenstein, George and O'Donoghue, Ted (2002), "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, vol XL, pp. 351-401

[34] Harris, Christopher and Laibson, David, (2004), "Instantaneous Gratification," *Manuscript*

[35] Heath, Chip, and Soll, Jack B., (1996), "Mental Budgeting and Consumer Decisions," *Journal of Consumer Research*, vol 23, pp. 40-52

[36] Kahneman, D., Wakker, P. & Sarin, R., (1997), "Back to Bentham? Explorations of experienced utility," *The Quarterly Journal of Economics*, 112, pp. 375-406

[37] Kim, Jeong-Yoo, (2006), "Hyperbolic Discounting and the Repeated Self-Control Problem," *Journal of Economic Psychology*, vol 27, pp. 344-359

[38] Krusell, P. and Smith, A. A. (1998), "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, pp. 867-896

[39] Laibson, David, (1997), "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, pp.443-477

[40] Lowenstein, George, (1996), "Out of Control: Visceral Influences on Behaviours," *Organizational Behaviour and Human Decision Processes* 65, pp.272-292

[41] Loewenstein, George, (1999), "Experimental Economics from the Vantage-Point of Behavioural Economics," *Economic Journal*, Royal Economic Society, vol. 109(453), pp. F23-34

[42] Loewenstein, George and David Schkade, (1999), "Wouldn't it be nice? Predicting future feelings," in *Well-Being: The Foundations of Hedonic Psychology*, edited by David Kahneman, Ed Diener and Norbert Schwarz, New York: Sage Found

[43] Mailath, George J., (1987), "Incentive Compatibility in Signaling Games with a Continuum of Types," *Econometrica*, vol 55(6), pp. 1349-1365

[44] Mulligan, Casey (1997), Parental Priorities and Economic Inequality, Chicago: University of Chicago Press

[45] de Nardi, Mariacristina, (2004), "Wealth, Inequality and Intergenerational Links," *Review of Economic Studies,* vol 71, pp. 743-768

[46] O'Donoghue, Ted and Matthew Rabin, (1999), "Doing It Now or Later," *American Economic Review*, 89, pp. 103-124

[47] O'Donoghue, Ted and Matthew Rabin, (2002), " Addiction and Present-Biased Preferences," *UCB Dept of Economics, paper E02'312*

[48] O'Donoghue, Ted and Matthew Rabin, (2006), "Incentives and Self Control," *UCLA Department of Economics Discussion Paper*, number 122247000000001262

[49] Orphanides, Athanasios and David Zervos, (1995), "Rational Addiction with Learning and Regret," *Journal of Political Economy*, 103, pp. 739-758

[50] Solon, G., "Mobility within and between generations," *mimeo*

[51] Spence, Michael, (1973), "Job Market Signaling," *The Quarterly Journal of Economics*, vol 87(3), pp. 355-374

[52] Thaler, Richard, (1980), "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behaviour and Organisation* vol 1(1), pp. 39-60

[53] Thaler Richard, (1981), "Some Empirical Evidence on Dynamic Inconsistency," *Economic Letters*, vol 8(3), pp. 201-207

[54] Thaler, Richard, (1985), "Mental Accounting and Consumer Choice," *Marketing Sci*, 4, pp. 199-214

[55] Thaler, Richard and Hersch M. Schefrin, (1981), "An Economic Theory of Self Control," *Journal of Political Economy*, 89, pp. 392-406

[56] Tournemaine, Frederic and Tsoukis, Christopher, (2007), "Status, Fertility, Growth and the Great Transition," *MPRA Paper number 8669.*

[57] Zelizer, Vivian, (1997), *The Social Meaning of Money*, Princeton, NJ, Princeton University Press